LNCS 6764

Julie A. Jacko (Ed.)

Human-Computer Interaction

Users and Applications

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

Part IV

HCI2011 INTERNATIONAL

Springer

Lecture Notes in Computer Science 6764

Julie A. Jacko (Ed.)

Human-Computer Interaction

Users and Applications

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

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 978-3-642-21618-3 e-ISBN 978-3-642-21619-0
DOI 10.1007/978-3-642-21619-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929076

CR Subject Classification (1998): H.5.2, H.5, H.2-4, I.2.10, I.4, J.1

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:

- HCI and learning
- Health and medicine applications
- Business and commerce
- HCI in complex environments
- Design and usability case studies
- Children and HCI
- Playing experience

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 3, LNCS 6763, Human–Computer Interaction—Towards Mobile and Intelligent Interaction Environments (Part III), 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 IV

Part I: HCI and Learning

Part II: Health and Medicine Applications

Part III: Business and Commerce

Part IV: HCI in Complex Environments

Part V: Design and Usability Case Studies

Part VI: Children and HCI

Part VII: Playing Experience

Part I

HCI and Learning

A Web-Based Learning Environment to Support Chemistry

Candice Adams and Cheryl Seals

Auburn University, Department of Computer Science and Software Engineering, Suite 3101 Shelby Center, Auburn, AL 36830 USA
{candice.adams,sealscd}@auburn.edu

Abstract. Increasing the use of technology is essential for the vitality of today's classroom because technology creates an environment that fosters positive attitudes toward learning and increase technologically literacy among K-12 students. Web-based applications that are centered on classroom lessons are effective at increasing literacy because it heightens students' motivation and interaction with technology. This paper will address the effect technology can have on advancing computer literacy when incorporated into the classroom through web-based learning environments. Moreover, it will take you behind the scenes from design to implementation of ChemiNet which is a web-based learning environment to support introductory Chemistry.

Keywords: Web-based Learning Environment (WLE), Human Computer Interaction, e-Learning, Instructional Design.

1 Introduction

In these economic challenging times, many states are cutting back on educational funding due to massive budget cuts. These budget cuts threaten educational leaders and school districts ability to implement new technologies, raise the quality of instruction in the classroom, and minimize achievement gaps among students from different backgrounds. According to a recent survey done by the American Association of School Administrators, "fifty-seven percent of respondents [state leaders] said they plan to delay or eliminate instructional improvements [technology] purchases in 2010-2011" [10] because of significant unmet needs in other areas. In order to lessen the effect of budget cuts on the quality of education being rendered by teachers to the students, there is a serious need to support classrooms instruction by introducing web-based education software in place of traditional educational software media.

Web-based educational support resources allow teachers to supplement their lesson plans by allowing their students to explore more concepts using the Internet. With increases in shipping, labor, raw materials, and other over head cost most software companies have to bare in order to produce quality software, it makes the products very expensive and unaffordable to many school districts across the country. However, web-resources allow teachers to maximize their classroom budget by

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 3–11, 2011.

utilizing resources they already have in the classroom, such as desktop and laptop computers, and Internet access. Teachers are embracing web-based instruction applications because it allows students to continue their study at home if they have access to a computer and the Internet, which most households equipped with these items. In 2003, a study was done by the U.S. Department of Commerce along with the U.S Census Bureau that revealed "majority of households have personal computers and Internet access" [12]. As documented in the October 2005 release of this report, almost 62% of households have computers and approx 55% have access to the Internet [12].

Teachers are encouraged to use classroom computers to as a supplement to traditional instruction methods. However, many of the computers are being used very limited to lack of relevant software that fits into their already designed lessons. In this regard, computers are being used for more casual purposes such as browsing the Internet or playing non-educational computer games. With technology expanding and the affordability of computers, computers used for instructional purposes have increased since 1983; from under 1 million to over 8 million, see figure 1.

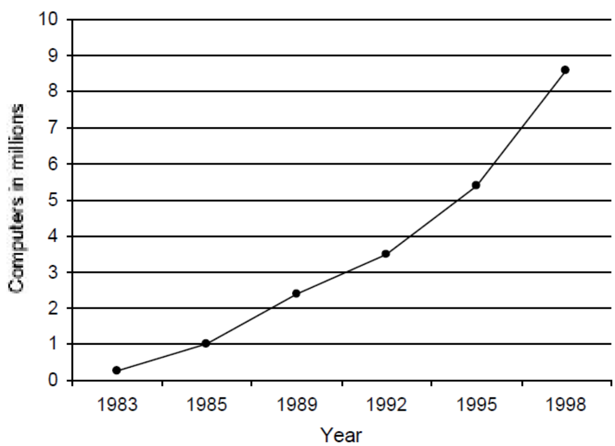


Fig. 1. Total Classroom Computers in the USA

With public schools in the United States already spending a significant amount of their budget on purchasing and maintaining computer hardware, there is little left to purchase software applications or additional equipment to run the software application. Moreover, as of fall 2005, close to 100% of all public schools across the United States has access to the Internet [8]. Many web applications can be found for free or diminutive fee compared to windows applications or applications that require a particular device in order to be effective in raising the quality of education in the classroom. A spokesman for SAS Curriculum Pathways, which provides web-based lessons and activities in core subjects, states that even a small "minimal charge was a barrier to adoption - so we removed it" [9]. For example, many novice developers have taken on freelance projects developing applications for the iPhone, iPad, or SMART-table. However, these devices cause several thousands of dollars to

Fig. 2. Students using SMART-Table

purchase for an average classroom size of fifteen to twenty-five students. If a class is able to purchase for example one SMART-table, only four to five students can utilize the table at once. Therefore, the students that are waiting their turn will become distracted or bored from the assignments they are working on while other students are participate in fun interactive activities using the SMART-board (see Figure 2 below).

Due to this limitation and the price of this new technology, it is more budget friendly for school districts seek out free or low cost web-based instructional support software.

1.1 Need

For years now, the talk of expanding the use of technology into the classroom has been a number one priority for many school districts. However, due to budget cuts and spending limits put on purchasing software and new hardware equipment there is a need for more affordable options for educational and instructional support software. Many school districts are placing huge expectations on teachers, however, if the money or effective software that can be incorporated with current lesson plans is not available, then teachers are being place at a disadvantage.

2 Educational Software

In today's thriving technological age, many secondary school systems are purchasing educational software to use as a part of their classroom instruction. The software may serve as an introduction to a lecture the instructor plans to teach, the actual lesson or a post-review of a lesson the instructor has already taught, but will reinforce some of the concepts by providing additional examples and a chance for students to explore the concepts on their own. Therefore, many secondary schools are integrating technology into the classroom each year by placing smart boards which are interactive white boards or additional computers terminals inside the classroom. They are also attending educational conference and events to seek out new software that they can use in the classroom and the most effect way to train teachers to use the software because teachers must be trained in how to incorporate the new resources into their everyday classroom activities for the software to be effective on the students achievements. Local community leaders are setting goals to increase the number of computers in the classroom in their school district with access to the Internet [2]. With additional funding and classroom-ready resources, there is a need for the development of educational software that is adequate for today's generation of secondary school-age children.

Educational software provides a collaborative dynamic learning experience for students to improve their knowledge in the software target area of study. It also provides a self-learning and adjustable pace environment for students to further explore and investigate topics discussed in the classroom. Therefore, educational software serves as a tool to support the material that is already being present in a traditional format by the teacher. Many children become easily distracted or disinterested when teachers are presenting material in the traditional way. Therefore, it is important to integrate technology into the lessons that are being taught because it can act as a "hook that gets students to participate" [7]. For example, if a teacher is presenting a lot of different kinds of elements from the periodic table on the board through a regular overhead projector, most students will become very overwhelm and stop absorbing the information. However, if you had a lesson on the computer where students can click and interact with the different elements and control the amount of information that take in at once, it will provide them with a better learning experience. Every educational application may not be effective for each class; therefore, it is the teacher's responsibility to request the most effective software that meets the instructional goals for the course.

2.1 Web-Based Tools and Technologies

With the popularity of the internet, web applications are becoming more standard because of the ubiquity of web browsers. Web applications can be access from many portable devices such as smart-phones, Netbooks, laptops, PDA's as long as there is a network connection or from non-portable devices at home, office, or a local library. More and more restaurants and shops are offering free or low-cost Wi-Fi connections to allow people with portable devices to connect while shopping or eating.

Computer software application manufactures are choosing to develop web applications over traditional windows applications because it allows them to update and maintain the applications much easier. If they built and sold windows applications, if a major update or defect arises they would have to distribute or install the new version of the application on potentially thousands of client computers, instead of just updating a single server with a new version of the application.

Many programming languages and software packages are used to build educational websites in today's technological advancing world. Some of these programming languages and software packages include JAVA, Ruby on Rails, C#, Visual Basic, Adobe Dreamweaver, Microsoft's ExpressionWeb, and several others.

There are also several new technologies that have been developed to make web application a more interactive experience. Some of these technologies are JavaScript, Ajax, Flash, and Silverlight. These new technologies allow you to provide a full interactive user experience to the user.

3 ChemiNet

Using an agile software development process an interactive web-based learning environment was developed to provide an exceptional user experience to middle and high school chemistry students. This new web-based learning environment is called ChemiNet, shown below in Figure 3. ChemiNet is designed to be a web-based learning environment where students can go through chemistry lessons as well as take quizzes based on a particular lesson. The application also features a fully interactive periodic table that allows students to go through and learn about each element. ChemiNet is a web-based application which means it can be access from any computer with Internet access.

By choosing the .NET framework, a web platform, to implement ChemiNet makes it so versatile. Students can have access to ChemiNet from any computer with Internet access whether at school or at home. Because students can work collaboratively or independently, ChemiNet was designed to be easy to use with the student's understanding of the material being taught as the ultimate goal of the application.

The bright colors used throughout the ChemiNet application on the user interfaces takes into to consideration the audience for this application. Secondary Education students tend to like things with bright colors and comical text. Also, the center of the home screen interface [see figure 3] contains a fun fact that changes when the screen refreshes or the page reloads. By offering such fun facts, students are more motivated to visit the site so they can find out another fact about Chemistry to share with their classmates or family.

Fig. 3. ChemiNet Home Screen

The goal of ChemiNet is to make the user interface easy to use, easy to learn and as less confusing as possible so that students using the application can have a pleasant learning experience. As shown in figure 3, you can see some of the design choices that were made in order to satisfy ChemiNet goal.

3.1 Home Splash Screen

With any web-based learning environment, it is necessary to pay close attention to the design of the user interface. From the home screen of the ChemiNet application, a user can navigate to any part of the application by using the non-intrusive navigation at the top and/or left side of the screen. One design decision that was made was to have modules, so a limited the number of options on the home screen. Based on this decision, larger buttons were chose for these options. The large vertical menu buttons on the left side of the screen makes it easy for students to click without mistakenly clicking the wrong option, see figure 3.

The vertical menu section consists of four options. These options are Lessons, Quizzes, Common Formulas, and Periodic Table. Once the user moves the mouse over an option, the text turns to yellow to notify the user that this can be selected. This ensures the user is aware that this option is clickable. The horizontal menu includes more of the standard options a user will normally see if other web applications. These options are home, help, and about. This is done because it will enable the user to return to home or get help at anytime while accessing the application. Therefore, each screen has a consistent screen layout in terms of the presence and location horizontal menu area. However, the vertical menu only shows on the home screen to continue with the non-instructive navigation scheme. The accessibility of the ChemiNet application is shown in figure 4.

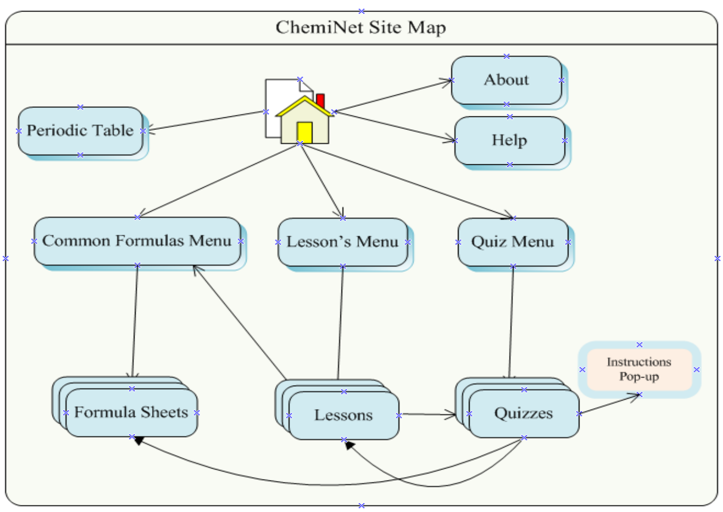


Fig. 4. ChemiNet Site Map

3.2 Module Screens

There are four modules in the ChemiNet application. The material within the application is broken down this way that students will be able to quickly and easily find the section they are seeking within the application. For example, the quiz and lesson modules are separate because if a student finished a particular lesson, but did not have the time to take the quiz associate with that lesson, they can do that by directly accessing the quiz module and finding the quiz for that particular lesson. They do not need to go through the lesson again in order to access the quiz. So, in order to deliver an effective learning experience each aspect of the application was carefully positioned so little confusion and frustration will occur while using the application. Two of the four modules will be discussed.

Lesson Module. The lessons module is the first option in the vertical menu on the home screen. When you click on the lesson's module, it transfers you to the screen as shown below in figure 5. Due to the time constraints of this project, there is currently only one chapter that contains three lessons. Once more chapters with lessons are implemented, a dropdown list will appear and allow the user to choose which subject area or chapter they are seeking lessons for within the lesson's module. Noticed the vertical menu has been removed to allow additional room on the page. The goal was to make the page size as small as possible, so that if a student has a smaller screen, they will not have to scroll horizontally to see the full page. According to Jakob Nielsen, "avoiding horizontal scrolling" is an essential usability guideline [13].

Periodic Table Module. The periodic table module contains a fully interactive periodic table, see figure 6 below. The goal of this module is to allow student to have fun while quickly finding information on a particular element. By position the mouse over in any element, a student can immediately see additional information pertaining to that element.

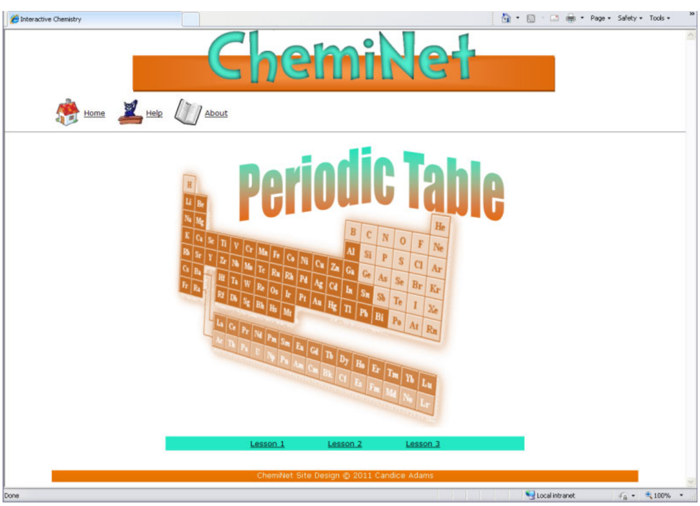


Fig. 5. Lesson Module

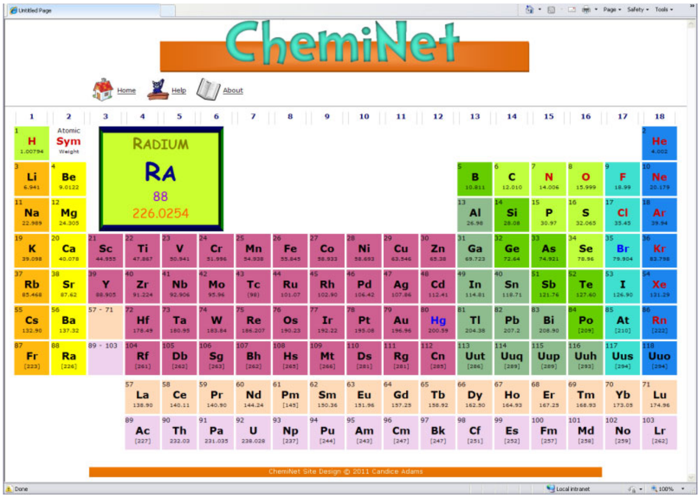


Fig. 6. Periodic Table Module

4 Results and Future Work

We have completed system requirements, system design, and development of ChemiNet. Our next step in this project is to perform usability analysis with K-12 and introductory level college students in first courses of chemistry. This will provide valuable feedback on the usability and usefulness of this project. Our hope with this work is to create an online learning tool that will support a virtual community of chemistry students and their teachers. This initial prototype will provide valuable feedback to improve this tool and to provide even greater support for chemistry teachers and students in the future.

References

1. Brusilovsky, P.: Adaptive Educational Systems on the World Wide Web: A review of available technologies. In: Proceedings of the Workshop "WWW-Based Tutoring" at the 4th International Conference on Intelligent Tutoring Systems, ITS 1998 (1998)
2. Stahl, G., Koschmann, T., Suthers, D.: Computer-supported collaborative learning: An historical perspective. In: Cambridge Handbook of the Learning Sciences. Cambridge University Press, Cambridge (2006)
3. Furner, J.M., Daigle, D.: The educational software/website effectiveness survey. International Journal of Instructional Media 31, 61–77 (2004)
4. Johnson, J.M.: Then, Now and Beyond: From lofty beginnings to the age of accountability. In: Learning and Leading with Technology, vol. 30(7). International Society for Technology in Education (2003)
5. Chappell, D..: Introducing the .NET Framework 3.5. Chappell & Associates, Microsoft Corporation (2007)
6. Dix, A., Finlay, J., Abowd, G.D., Beale, R.: Human-Computer Interaction, 3rd edn. Prentice Hall, Englewood Cliffs (2004)

7. Blumenfeld, P.C., Kempler, T.M., Krajcik, J.S.: Motivation and Cognitive Engagement in Learning Environments. In: Sawyer, R.K. (ed.) The Cambridge Handbook of the Learning Sciences. New York, Cambridge (2006)
8. U.S. Department of Education, National Center for Education Statistics.: Internet Access in U.S. Public Schools and Classrooms: 1994-2005 (NCES 2007-020) (2007)
9. Prabhu, M.T.: Free online curriculum expanding to middle grades. eSchoolNews (2010), `http://www.eschoolnews.com/2010/05/06/free-online-curriculum-expanding-to-middlegrades/`
10. eSchoolNews: Survey: School budget cuts even worse next year (2010), `https://www.eschoolnews.com/2010/04/09/survey-school-budget-cuts-even-worst-next-year/`
11. eSchoolNews.: Free tool lets students participate during class (2010), `http://www.eschoolnews.com/2010/04/30/free-tool-lets-students-participate-during-class/`
12. U.S. Department of Commerce Economics and Statistics Administration. U.S. Census Bureau.: Computer and Internet Use in the United States: 2003. Issued October 2005 P203- 208 (2005)
13. Nielsen, J.: Scrolling and Scrollbars (2005), `http://www.useit.com/alertbox/20050711.html`

Introducing Mobility in Serious Games: Enhancing Situated and Collaborative Learning

Sébastien George and Audrey Serna

Université de Lyon, INSA-Lyon, LIESP Laboratory, F-69621, Villeurbanne, France
{Sebastien.George,Audrey.Serna}@insa-lyon.fr

Abstract. This paper explores the introduction of mobile technologies in a particular serious games subset called learning games. We focus on two main purposes where mobility turns out to be useful: situated learning and collaborative learning. The article outlines the opportunities and the remaining challenges for these learning situations. Some important issues are highlighted for immerging learners in real context. We also propose some scenarios to illustrate the benefits of mobile devices in classroom situations.

Keywords: Mobility, learning games, situated learning, collaborative learning, outdoor mobile learning, in-class mobile learning.

1 Introduction

Mobile devices are increasingly present in our professional and personal lives. Their growing capabilities (memory, CPU), their connectivity (3G, wireless technologies) and their features (web browser, GPS, camera and video recording. etc.) open new opportunities, particularly in the field of education. Mobile learning brings new technological and educational challenges as evidenced by the many issues raised during recent conferences on this topic. For instance, mobile technologies enable more motivating learning experiences.

In this paper, the potential of the use of mobile technologies in serious games is explored. Serious games are "*games used for purposes other than mere entertainment*" [1]. We focus specifically on learning games, a serious game category aiming at using game mechanisms to promote the learning of knowledge and the building of skills.

Mobility and mobile technologies could be of a great interest in learning games situations for two main purposes:

- to favor situated learning,
- to promote interaction and collaboration between learners learners for in-class situations.

2 Mobility Technologies to Favor Situated Learning with Games

Mobile learning, or M-learning, allows learners to access learning material anytime and anywhere through several devices. From our point of view, a M-learning

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 12–20, 2011.

application must be able to extract, interpret and use contextual information and adapt functionalities to the current context of use. In this way, the learning could take place in situation. We refer to situated learning theory [2]. In this theory, learning act is situated in the activity in which it takes place. Situated learning occurs when learners work on authentic tasks that take place in real world. Using learning games based on role-playing in a real situation could favor the learning of particular behaviors. Learning may occur in location and time which are significant and relevant for learners.

Several innovative educational situations can be identified. We illustrate them in the following parts of this section.

2.1 Taking the Natural Environment into Consideration

A possibility with mobile learning is to use the natural environment as a source of information. Some knowledge needs students learning through observation and is not very easy to teach by either traditional classroom teaching or web-based learning environment. Mobile application proves to be useful in this case. For example, the “Butterfly-Watching System” [3] supports an activity in which each learner takes a butterfly picture with a PDA (Fig. 1). Retrieval technique is applied on a database to search for the most closely matching butterfly information and is returned in real time to the learner’s device.

Fig. 1. The Butterfly-Watching System [1]

With a similar idea based on observation, an augmented reality game was designed to teach zoo visitors about the illegal wildlife trade [4]. So natural science learning could benefit from outdoor mobile learning game technologies by allowing open activities as scavenger hunt or identification in a natural environment.

2.2 Engaging Learners into Simulation as Part of a Dynamic System

Mobile learning offers the opportunity to extend the notion of microworld [5]. In a microworld, learners can explore alternatives, test hypotheses, and discover facts about that world. With mobile technologies, the learners themselves can act in an immersive simulation of a dynamic system. For instance, in Virus Game [6], learners took part in a participatory simulation about the spread of a virus. Students wear an active badge and the proximity with another student may spread the virus. Learners can define strategies to understand how the virus is spreading, who is the initial infected person, who is immune… The result is that students were really engaged with the simulation, and found it to be a rewarding and stimulating experience.

Another example of learning game engaging the students into a simulation is Savannah [7]. In this game, children use PDAs, moving around a playing field outdoors and acting as lions. The main challenge is to understand and survive in a territory. They have to collaborate in order to achieve the games objectives.

The goal of these kinds of research works is to move a simulation into the real world, so students can interact with and see the effects immediately. By making them actors of the simulation, they are more engaged in the learning process. Naismith *et al.* [8] highlight the benefit of having a learner, through a networked device, become part of a dynamic system: “they do not just watch the simulation, they are the simulation”.

2.3 Favoring the Learning of Professional Skills and Gestures

Combining the real and the virtual can be a good solution to achieve educational goals. Augmented reality techniques could be used to teach some gestures. For instance, HMTD (Help Me To Do) [9] uses wearable computer and head-mounted display to teach maintenance and repairing of professional appliances. The main idea is to put the user in a precise situation (use, maintenance, diagnostic or repairing) to to understand functioning principles and commands. The augmented reality allows to link different representation of the same information.

The Context-Aware Agent-Supported Augmented Reality System (CAARS) [10] is an industry research initiative developed for the manufacturing training domain. This

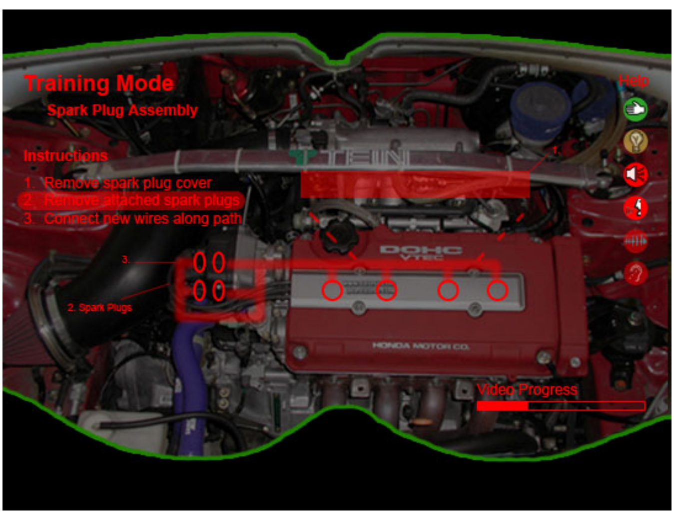


Fig. 2. CAARS user’s view of an engine with see-through goggles [10]

system has been extended for developing a mobile augmented reality learning game with a virtual instructor in mixed reality. The goal is to favor the learning of professional skills as well as psychomotor task in real world environments.

This kind of mobile augmented reality learning game is at the beginning. Learning scenarios are quite basic at the moment. We think that next generation will integrate more elaborate and collaborative games scenarios that will increase the learners' motivation.

2.4 Mobile Learning Game for Situated Learning: Main Issues

Mobile learning game still in its infancy and a lot of research works should be done for a deeper understanding of its design principle and learning impacts. We can highlight several issues:

- **Supporting temporal mobility:** Dispersion in time makes it hard to define precisely the start and end of a mobile learning session. Learning scenario should consider this aspect and support not only the spatial mobility of students, but also their temporal mobility (for example by giving the opportunity to easily stop and resume a session). Facilities should also be provided to facilitate asynchronous discussions in a contextual way [11].
- **Integrating mobile applications into global learning systems:** The question of the integration of a mobile learning game application can be addressed from two points of view. First, a learner should be able to switch seemlessly between a desktop application and a mobile application. This HCI concept can be done with solutions of user interfaces redistribution [12]. Secondly, information gathered during mobile use could be reuse on desktop applications in order to adapt the learning game to each learner experience. So, mobile learning applications should be seen as a part of a whole, in relation to other learning tools used by students.
- **Making a room for the teacher:** we know that human teachers play an important role in e-learning, even if virtual tutors are used. Teachers favor the transfer of knowledge from one situation to another, from one context to another. So the role of teacher should be deeper integrated in mobile learning game scenarios. In particular, tutoring should be facilitated by giving teachers specific tools to finely monitor students' activities.

To sum up, in situated learning research, mobility is considered as a key issue for immerging learners in real context. From this perspective, most of mobile learning games are designed for "outside the classroom" scenarios. However, introducing mobility in a well-known and static environment, such as the classroom, can also present some advantages.

3 Introducing Mobile Devices into the Classroom to Promote Learners Interaction

The introduction of mobile technologies into the classroom can present some advantages in terms of collaborative learning. This theory is based on the role of social interactions in the process of learning. Mobile devices present two advantages that seem to favor social interactions and collaborative situations. First, they can easily communicate with other devices. Secondly, collaboration between learners is possible through and around the device [8].

3.1 Existing Systems Using Mobility in the Classroom

Concretely, a major part of the studies focusing on the introduction of mobile devices into classrooms are limited to systems that gather responses and provide feedback during in-class exercises. Several studies use mobile phones for submitting solutions in the form of text or photos via SMS/MMS services during exercises or revisions [13], [14]. Experimenting these systems underline usage limitations due to the device features. Learners regret, for instance, a lack of detailed feedback due to the limited size of the screen, and they express a desire to use other platforms to access to more functions. They suggest, for example, to use laptops or tablets to have a best vision of their work or to consult resources such as Wikipedia. To overcome partially these problems, others studies prefer the use of tablet-PC to allow learners to submit digital ink answers [15]. However, all these studies are based on a behaviorist learning approach: they focus on learners' personal work and are applied to classic exercises, neither game motives nor collaboration interactions are used.

In the end, to come closer to the usage of mobile technologies that we expect into the classroom, research on mobile-CSCL (computer-supported collaborative learning) presents some interesting leads. MCSCL explore the use of handheld devices to support a natural mobile collaboration environment with face-to-face social interactions [16]. In particular, two studies catch our attention. The first one introduces handheld devices in a collaborative learning (CL) activity for ordering numbers for children [16]. The authors use mobility to overcome usability problems encountered in the classic CL activity. In particular, mobility allows enhancing children' participation and social interactions, including coordination, organization of individual work with joint group activities, involvement in the activity, negotiation, synchronization, etc. The second significant work, called Caretta, introduces mixed reality into the classroom to favor face-to-face collaboration [17]. A sensing board with RFID recognition insures collaboration in a shared space via physical objects and PDA devices support simulation activity in personal spaces. Mobility is here considered as a support for personal reflection within a CL activity.

3.2 Main Issues to Consider When Using Mobility in the Classroom

This brief overview of literature demonstrates that the introduction of mobile technology in the classroom is not fully exploited and points out some considerations to be taken into account when designing mobile learning games in the classroom:

- Whatever the approach chosen (behaviorist or collaborative learning), the use of mobile devices has to be combined with others devices. In the MCSCM approach, the combination of (homogeneous or heterogeneous) platforms allows personal work as well as synchronization/collaboration with the rest of the group. As counterexample, the exclusive use of mobile phones for in-class exercises revealed lack of significant detailed information and too restricted functionalities.

- Mobile technologies can strengthen the success of learning games. Schwabe & Göth [18] address some general issues in designing mobile learning games in regard to social aspects (such as face-to-face collaborations) and representation of interesting aspects of reality (with augmented reality for instance), leading to better focus and motivation from learners.
  Despite these advantages, the use of mobile devices inside the classroom is little explored in mobile serious games (except in the MCSCL approach).
- The studies presented here are considering mobility from the learners' perspective. Very few studies envisage teachers as end-users with mobile tutoring requirements. So far, tools developed for teachers using mobile technologies concentrate on management of administrative tasks [8].

3.3 Illustration of Mobility Benefits in the Classroom

From these considerations, different contexts of use emerge from the introduction of mobile device into a classroom. For example, learners could use a platforms composition (using a shared computer and their personal Smartphone), or they can play different roles (like in Savannah [7]). To be efficient, the interface of the learning activity (or educational scenario) must support and must be adapted to the different potential contexts of use. To go further, we propose to consider the introduction of mobile technologies on learning games relying on a HCI concept called plasticity. Plasticity denotes the capacity of user interfaces (UI) to adapt to the variation of context of use while preserving usability [19]. Depending on the situation, UI can be transformed (remolding) or can be redistributed among the available interaction resources (redistribution by partial or total migration [12]). More than adaptation itself, we'll use these mechanisms to explore the possible usage of mobile devices into learning games (see [20] for similar approach for e-government).

In a classic learning game session held on our industrial engineering school department, students are working by groups. They usually follow the educational scenario on a same computer and collaborative tasks are done with pen and paper. During a session, the time is shared between group work and debriefing with the tutor. For example, in the game named "Puissance 7" (P7), students have to solve an industrial problem by choosing well-known quality tools. In this particular context, we identify several situations where the use of mobile devices combined with plasticity mechanisms can enhance collaborative learning.

Moment 1
John, Helena and Peter are taking part of P7. In the educational scenario, they have to lead an investigation to find out the causes of a problem in a particular company and choose solutions (learning problem solving quality methodology and tools) on a shared computer. At each stage of the game, they have to choose the appropriate tool from the quality methodology. They can vote and justify their choices on their Smartphone. Peter doesn't choose the same tool as John and Helena. After discussion, he let be persuaded and they validate their final choice on the shared computer so that they can reach the next stage of the game.

Analysis

Thanks to a distribution of the HCI on both personal Smartphone's and shared computer, learners can take their own decisions and share or synchronize them with the rest of the group. Therefore, the platforms composition guarantees the regulation of social interactions offering decision making and negotiation spaces. In addition, as in Zurita *et al.* [16] study, the introduction of mobile devices assures that each member of the group participated.

Moment 2
As they advanced in the investigation, John, Helena and Peter collect information by interrogating people in different departments of the simulated company. But Helena is a little confused with all the information collected. Whereas her friends are completing the elements that they found on the shared computer, she decides to look back to what they discovered. She uses her Smartphone to look at the company printed map as if she had a magnifying glass: thanks to the augmented system she can easily repair the departments already visited (additional information is added to the camera-based vision on the screen of the Smartphone). She comes closer to the technical department and the information collected is displayed. She can play again the interview with the technical manager, check the cues collected, etc. She can specify the importance of the information. By moving away her Smartphone, she comes back to the global augmented vision of the company map. The technical department is now marked as containing major information.

Analysis

The augmented system allows each learner to trace the investigation from a personal perspective (adding some personal assessment on collected information for instance). As suggested in Caretta [17], the use of Smartphones creates personal spaces that support individual activity without interfering with group activities (on the shared computer). In addition, the use of augmented reality should enhance the immersion of learners into the game leading them to get more involved in the investigation (as in [18]).

Moment 3
At the end of a game level (or stage), John, Helena and Peter meet again Patrick, the SG tutor, and the rest of the class at the central table for a debriefing session. John put his Smartphone on the central table. Doing so, the HCI offers him a global vision of the problem solving methodology and the course of his own team. As Patrick is explaining some theoretical concepts manipulated during the previous stage, John can inspect on his Smartphone all the tools used previously and the consequences on their investigation and on the decisions they made.

Analysis

During collective sessions (moment 1 and moment 2), tools provided on mobile devices allow learners to work individually within collaborative activities and they are combined to the shared platform for coordination/synchronization. In these sessions, learners are playing the role of quality consultants and are part of team. When they come to the central table for debriefing sessions, they occupy the role of student reflecting on the methodology learned, tools choices, consequences of a decision, etc. (the concept of multirole support is explored in Savannah [7]). Regarding HCI plasticity, the change of role triggers an adaptation the Smartphone HCI that will provide appropriated tools.

Moment 4
Patrick is the tutor of P7. During collaborative sessions, Patrick has to help the different groups in their investigation giving them some clues. Thanks to his tablet, he can visualize additional information regarding the investigation (such as recognized key-words in the interviews phases). In addition, his tablet provides him an overview of groups' progress in order to help him to synchronize the different groups when needed. When he gets closer to a particular group, Patrick can migrate the detailed view of the group progress on his phone. He can also trigger the synchronization by using his tablet as a remote control (the game level will be automatically completed).

Analysis

Here the mobile device is considered from the tutor or teacher perspective. The functionalities and information provided by the system are adapted according to the position of the tutor within the classroom (his proximity to a particular group for instance).

Future work will be focused on the implementation of presented scenarios for collaborative learning enhancement and validation with users' experimentation.

4 Conclusion

In this paper, we present how the use of mobile technologies in learning games can favor situated learning and collaborative learning. We outlined the opportunities and the remaining challenges for each of these learning situations. We have also identified common issues which are important to address for the future of mobile game-based learning. A first aspect concerns authoring tools. The design and development of mobile learning games is complex and time consuming. Methods and tools currently exist to help the design of learning games [21]. New functionalities should be added to give authors the possibility of exploring pedagogical innovations with mobile devices. A second issue is related to the place of the teacher in mobile learning game situations. For both classroom and outdoor contexts, learning impacts greatly depends on the quality of human tutoring. The design of a specific and adapted instrumentation is essential in order to support this tutoring activity in a mobile learning environment.

References

1. Susi, T., Johannesson, M., Backlund, P.: Serious games – An overview. School of Humanities and Informatics. University of Skövde, Sweden (2007)
2. Lave, J., Wenger, E.: Situated Learning. Legitimate Peripheral Participation. Cambridge University Press, Cambridge (1991)
3. Chen, Y., Kao, T., Yu, G., Sheu, J.: A mobile butterfly-watching learning system for supporting independent learning. In: Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, JungLi, Taiwan, pp. 11–18 (2004)
4. Perry, J., Klopfer, E., Norton, M., Sutch, D., Sandford, R., Facer, K.: AR gone wild: two approaches to using augmented reality learning games in Zoos. In: George, S., Serna, A. (eds.) Proceedings of the 8th International Conference on International Conference for the

Learning Sciences, vol. 3, pp. 322–329. International Society of the Learning Sciences, Utrecht (2008)

5. Papert, S.: Mindstorms, Children. Computers and Powerful Ideas. Inc., New York (1980).
6. Colella, V.: Participatory Simulations: Building Collaborative Understanding Through Immersive Dynamic Modeling. Journal of the Learning Sciences 9, 471 (2000)
7. Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., Kirk, D.: Savannah: mobile gaming and learning. Journal of Computer Assisted Learning 20, 399–409 (2004)
8. Naismith, L., Lonsdale, P., Vavoula, G., Sharples, M.: Literature review in mobile technologies and learning. University of Birmingham (2004)
9. David, B., Chalon, R., Champalle, O., Masserey, G., Yin, C.: Contextual Mobile Learning: A Step Further to Mastering Professional Appliances. IEEE Multidisciplinary Engineering Education Magazine 2, 5–9 (2007)
10. Doswell, J., Harmeyer, K.: Extending the 'serious game'boundary: Virtual instructors in mobile mixed reality learning games. In: Digital Games Research Association International Conference (DiGRA 2007), Tokyo, Japan, pp. 524–529 (2007)
11. George, S., Lekira, A.: MeCoCo: A Context-Aware System for Mediated Communications. International Journal of Interactive Mobile Technologies (iJIM) 3, 26–33 (2009)
12. Balme, L., Demeure, A., Barralon, N., Calvary, G.: CAMELEON-RT: A software architecture reference model for distributed, migratable, and plastic user interfaces. In: Markopoulos, P., Eggen, B., Aarts, E., Crowley, J.L. (eds.) EUSAI 2004. LNCS, vol. 3295, pp. 291–302. Springer, Heidelberg (2004)
13. Lindquist, D., Denning, T., Kelly, M., Malani, R., Griswold, W.G., Simon, B.: Exploring the potential of mobile phones for active learning in the classroom. ACM SIGCSE Bulletin, 384–388 (2007)
14. Thornton, P., Houser, C.: Using Mobile Phones in Education. In: Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE 2004), pp. 3–10. IEEE Computer Society, Los Alamitos (2004)
15. Koile, K., Singer, D.: Improving learning in CS1 via tablet-PC-based in-class assessment. In: Proceedings of the Second International Workshop on Computing Education Research, pp. 119–126. ACM, New York (2006)
16. Zurita, G., Nussbaum, M., Shaples, M.: Encouraging Face-to-Face Collaborative Learning through the Use of Handheld Computers in the Classroom. In: Proceedings of the 5th International Symposium on Human Computer Interaction with Mobile Devices and Services, pp. 193–208. Springer, Udine (2003)
17. Sugimoto, M., Hosoi, K., Hiromichi, H.: Caretta: a system for supporting face-to-face collaboration by integrating personal and shared spaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 41–48. ACM, New York (2004)
18. Schwabe, G., Göth, C.: Mobile learning with a mobile game: design and motivational effects. Journal of Computer Assisted Learning 21, 204–216 (2005)
19. Thevenin, D., Coutaz, J.: Plasticity of user interfaces: Framework and research agenda. In: Proceedings of INTERACT 1999: IFIP TC. 13 International Conference on Human-Computer Interaction, Edinburgh, UK, pp. 110–117 (1999)
20. Calvary, G., Serna, A., Coutaz, J., Scapin, D., Pontico, F., Winckler, M.: Envisioning Advanced User Interfaces for E-Government Applications: A Case Study. In: Practical Studies in E-Government: Best Practices from Around the World, p. 205 (2010)
21. Marfisi-Schottman, I., George, S., Tarpin-Bernard, F.: Tools and Methods for Efficiently Designing Serious Games. In: Proceedings of the 4th Europeen Conference on Games Based Learning ECGBL 2010, Copenhagen, Denmark, pp. 226–234 (2010)

Visualization Framework for Computer System Learning

Eiichi Hayakawa[1], Yuuki Nakagawa[2], Hideharu Ochiai[2],
Masahiko Fuji[2], and Yosuke Nishino[2]

[1] Takushoku University, Faculty of Engineering,
Department of Computer Science
[2] Hachioji Soushi High-School
hayakawa@cs.takushoku-u.ac.jp,
{Yuuki.Nakagawa,Hideharu.Ochiai,Masahiki.Fuji,
Yosuke.Nishino}@os.cs.takushoku-u.ac.jp

Abstract. In this paper we describe a design of a computer system visualization framework for system software learning. It is difficult for learners to learn the behavior of computer system because of multiple layers of the system and invisibility. For the problem we developed and utilized an integrated learning environment. The results show the problem of high utilization cost of the environment and high development cost of the learning materials. We design a visualization framework based on cloud computing and a material development environment. The framework presents web based visualization component and log handling facilities. Lecturers can easily develop and provide the visualization material to utilize these facilities for building an appropriate material. In addition learner can use the visualization environment without installing because of running it on the cloud server. This leads to reduce the setup time for the lecturers.

1 Introduction

It is difficult for computer science learner to study the computer system such as hardware, OS and embedded systems. It is because the learner doesn't trace the behavior of the computer system easily to call each component asynchronously in the multilayer structure. Moreover, the behavior of the system might not effect the operation of the application directly. As a result, it is difficult for the learners to trace the behavior, as they don't understand to process the request from the application and the event from hardware.

To teach the system programming, a lot of teaching materials of making to visible are developed. In the embedded system, the robot device such as LEGO Mindstorms or line tracing cars and evaluation board are often used. Trace log from the devices visualizer [1] is developed for debugging. Our research group has also developed and utilized the integrated learning support environment named Minato [2]. In the Minato the internal behavior of OS and sensor data of the robot are logged and visualized as the state transition of OS processes and the track of sensor data.

In our experience it is apparent that the environment that easily assembles teaching materials is important. Rapid material development environment requires adapting

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 21–26, 2011.

various computer systems used in education. The materials easily adaptable to system software education are required because educational system target differs in the university, the enterprise or the carriculum.

In this report we describe a design and a prototype about visualization framework. We present various visualizing components for each layer of the system to assemble these components. Lecturers can create the material for their needs to build the visualization materials through Web browser. Learners can use these components that run on cloud servers in the web browser.

2 Requirement of the System

We consider that these requirements are important for system visualization environment through the Minato system development and the experiment in a high school.

(1) Customizability of materials
A development environment on which lecturer customizes materials is required. Our previous environment, Minato, integrates system log monitoring tool, OS or robot visualization tool and system simulator into one application. Learners can use many visualization materials without switching applications. Problems are, however, apparent that it is difficult to allow lecturers to customize the materials where they use OS or embedded system to their needs. In embedded system education every educational institution often utilizes different operating systems such as Android, TOPPERS or eCOS or embedded system such as ARM, MIPS or x86. Material builder requires the environment to easily build materials proper to their environment.

(2) Easily system installation
Easily installation and system management is also required. It is hard to prepare the visualization environment in Java because Java has different virtual machine versions and requires many libraries to install to the local machines. Since various machines such as Windows, Mac or Linux are set up in educational institution, learners are affected with the usability that the OS presented.

The target system requires that it is secure from learner's action and rapid preparation in the classroom. When executing exercise programs on real hardware, inappropriate application or system may crash the system itself and lose system log or tracing data. Moreover, it may be expensive to prepare real hardware to each learner in the classroom.

3 System Design

3.1 System Overall

Figure 1 shows the overall structure of the system. The system is built from three nodes: Component Node, Execution Node and Service Node.

First, Component Node presents the visualization materials. These are classified into visualizing component and combining component. Visualizing component is minimum component for system program visualization. Combining component is the

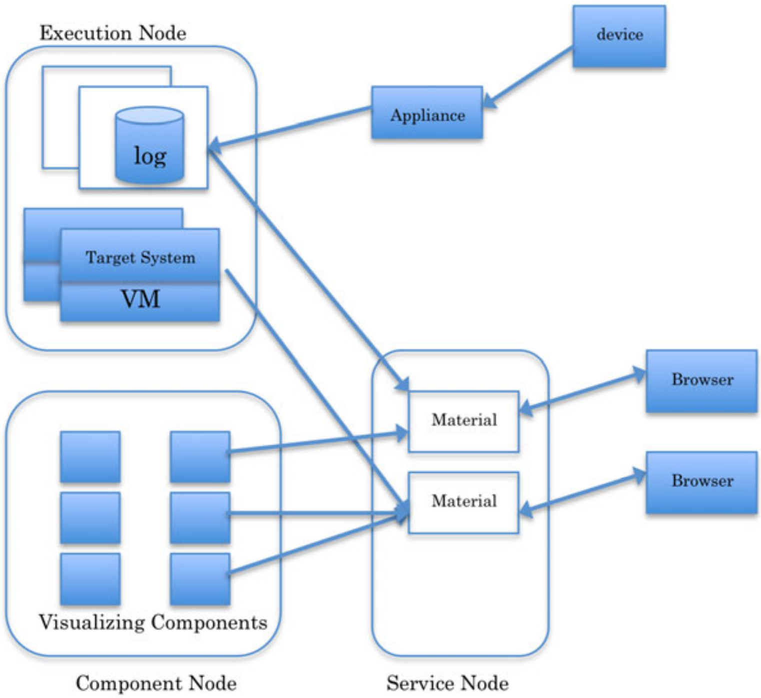


Fig. 1. Overall Structure of the System

aggregate of visualizing components. Lecturers build a material required as a combining component with visualizing components.

Second, Execution Node presents and controls an execution environment for a target system. The environment consists of virtual machine or real hardware system. This node acquires the execution information such as CPU performance, process creation and switch, memory usage and I/O and interrupt information from the target.

Third, Service Node presents materials as assembled components to the learner. This node runs on cloud server and maintains learner's information. Using cloud server, this service can scale out in proportion to increasing learners.

3.2 Component Node

Component Node implements in Javascript as jQuery[3] plugin. Common execution control methods such as to control animation speed, replay and step execution are presented as component common methods. The groups of visualization components are following:

- Electrocardiogram chart
- Event list
- State transition chart
- Sequential chart

- Performance chart
- Architectural chart
- Graph chart

Electrocardiogram chart displays system changes over time such as an electrocardiogram. Event list displays a list about timer or I/O events. Sequential chart also displays these events sequentially. State transition chart displays state transition diagram in the system's behavior such as process, file, etc. Performance chart displays line or bar graphs about system performance. Architectural chart displays hardware components such as CPU, memory, I/O devices. Graph chart displays tree or graph structure to represent the relation among system components.

Assembled component generates a visualization material to assemble visual components. For example, process visualization material is build to assemble Electrocardiogram chart and State transition chart. Electrocardiogram chart represents transitions about processes and State transition chart represents behavior about processes in the OS.

3.3 Execution Node

Execution Node is implemented based on KVM [4] kernel virtual machine or QEMU software emulator for hardware not to support KVM. Virtual machine is presented as a node controller of cloud servers to support the increase of users. Real hardware such as evaluation board or robot devices is connected through Execution Node from which components can acquire visualization data. We adopt common visualization data format as JSON because of affinity for Javascript.

Ftrace [5] used in Linux utilizes to acquire log data from OS. It is lower tracing overhead and effect to target system. System information such as process transition, interrupt or I/O handling and performance information are acquired through ftrace mechanism.

3.4 Cloud Server Platform

Service Node and Execution Node run on node controller in Eucalyptus that is open source cloud computing platform. We can present classroom size visualization materials to increase node controller executing Service Node and Execution Node. We don't require installing Java or visualization application on local machines but only require Web browser. Data from hardware such as robot transfers to the Execution Node through the medium of network appliance. It supports the control and log data transfer from the device when the hardware doesn't support TCP/IP network.

4 Prototype System

Figure 2 shows a prototype material for Android process visualization.

This educational material is build from three components. Upper half component visualizes state transitions about five processes. Lower left visualizes process queues in the Android OS. Green represents running process, yellow represents ready

Fig. 2. Prototype of Android visualization tool

processes and red represents suspend processes. Lower right visualizes the process tree of the Android. These components acquire visualization log data from Android simulator or board computer running Android in Execution Node.

Lecturers can add visualization materials not to rewrite the entire program when they require adding it. For example, they add the Javascript code to the assembled component that visualizes process list from log data using event list component when lecturers want to add active process list.

5 Conclusion

We describe the design of a computer system visualization framework for system software learning. The framework presents web based visualization materials with three nodes on cloud servers. Lecturers can easily develop and provide the visualization material to utilize these components for building an appropriate material. Running materials on cloud servers, learner is available to use them without installation. This leads to reduce the setup time for the lecturers.

Future work is completely implementing the system. Few components and materials are currently implemented. Previous programs used in the Minato is porting and rewriting to this system. After that we have plan to build the system learning materials using these components and evaluate for high school and university students.

References

1. Goto, J., Honda, S., Nagao, T., Takada, H.: Development of Visualization Tool for Trace Log, SIG Technical Reports, 2009-EMB-12, pp. 73–78 (2009) (in Japanese)
2. Nishino, Y., Hayakawa, E.: Minato: Integrated Visualization Environment for Embedded Systems Learning. LNCS, pp. 325–333 (2009)
3. jQuery, `http://www.jquery.com/`
4. KVM - Kernel Based Virtual Machine
5. `http://www.linux-kvm.org/page/Main_Page`
6. Ftrace Linux Kernel Tracing
7. `http://events.linuxfoundation.org/slides/2010/linuxcon_japan/linuxcon_jp2010_rostedt.pdf`
8. Eucalyptus Systems, `http://www.eucalyptus.com/`

Associating Learners' Cognitive Style with Their Navigation Behaviors: A Data-Mining Approach

Yung-Chi Hsu and Sherry Y. Chen

Graduate Institute of Network Learning Technology
National Central University, Taiwan
sherry@cl.ncu.edu.tw

Abstract. Investigating how learners' individual differences affect their navigation behavior can help us understand learners' preferences and can be used to develop the Web-based learning system that can meet learners' needs. Among various individual differences, learners' cognitive styles need to be considered because it is concerned how users process information. There is also a need to examine the relationship between their navigation behavior and performance because such findings can be used to suggest suitable navigation tools to meet learners' real needs. To this end, the aims of this study tend to use a data mining approach to not only investigate the relationship between learners' cognitive styles and their navigation behavior but also analyze how their navigation behavior influence performance. The results indicate that holists and serialists have different navigation behavior. However, there is no direct relationship between learning performance and navigation behavior.

Keyword: Data mining, cognitive styles, navigation behavior, learning performance.

1 Introduction

Web-based learning systems (WBLSs) become a popular topic recently. The reason is that such systems provide sufficient flexibilities, which allow learners to build their own navigation strategies (Wang et al., 2010). However, it may cause their diverse requirements when they navigate in the WBLSs. To address this issue, learners' individual differences are a key factor when developing such systems. For further investigating various individual differences, learners' cognitive styles may influence their navigation behavior (Clewley et al., 2010). For example, holists may prefer to use a navigation tool that can get an overall picture while serialists may prefer to use a tool that can get procedure details (Lee et al., 2005). In addition, recent studies also examined which navigation tools *are most suitable* to learners because the usage of navigation tools may influence their learning performance (Minetou et al., 2008).

To this end, our aims are two-folds: (a) to examine how learners' cognitive styles influence navigation behavior and (b) to examine learners' navigation behavior affects learning performance. Such investigation can contribute to the understanding of developing the systems that can meet diverse learners' needs. To reach our aims, a

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 27–34, 2011.

data-mining approach, especially association rules, is adopted because it is helpful to identify meaningful relationships between items that co-occur frequently. This paper is organized as follows. In section 2, the literature review is described. In section 3, the methodology used to achieve our aims is described. In section 4, the experimental results are presented. The conclusions are summarized in the last section.

2 Literature Review

Recently, several studies indicate that learners' cognitive styles are relevant to how learners prefer to organize and represent information. There are several dimensions of cognitive styles. Among them, serialists and holists have different characteristics (Pask, 1976). More specifically, holists prefer to process information in global ways while serialists prefer to build up procedural understanding step by step. In addition, Jonassen and Grabowski (1993) describe that holists prefer to focus on several aspects at the same time. They can have many goals and work on topics that span varying levels of structure. Conversely, serialists prefer to combine information in a linear way. In other words, they can only focus on a single concept at a time and work from the bottom up. Moreover, Ford (1993) indicated that holists strongly favored to use a tool which provides a global picture, such as a hierarchical map. In contrast, serialists prefer to use a tool which can provide procedure details, such as an alphabetical index. In summary, serialists and holists have different characteristics, which may influence their navigation strategies. However, such dimension is ignored in existing works (Pask, 1976). Thus, there is a need to consider such a dimension so that we can develop the WBLSs that can meet the diverse needs of holists and serialists.

Although the above studies have provided initial guidance for the design of navigation tools based on the needs of serialists and holists, they mainly use statistical analyses, which indicate some trends about learners' navigation behavior. However, it is difficult to find some unexpected relationships hidden in a dataset with the statistical techniques. To achieve this target, a data mining approach is more appropriate because it can be used to search hidden information from existing datasets (Li & Chen, 2009). To this end, our recent studies tend to use data mining approaches to evaluate the effects of cognitive styles. For example, Chen and Liu (2009) adopt a clustering method to investigate learners' learning behavior and performance in WBLS. Their finding indicates that Field Independent users frequently used an alphabetical index to locate the specific content while Field Dependent users usually chose a hierarchical map to obtain an overall picture. Such learning patterns also have great effects on their performance, especially for Field Dependent students. Moreover, Chen and Liu (2008) use both clustering and classification to analyze how cognitive styles affect students' navigation behavior. Their wok indicates that the same cognitive style group may achieve the same target with different ways, which reflect the characteristics of this type of cognitive style. These studies provide fruitful results, but they only focus on clustering and classification methods. Regarding the clustering, it is mainly used to distinguish similarities and differences among groups and identify a major trend (Han & Kamber, 2001). Regarding the classification, it is useful to decide how new items should be classified.

The aforementioned studies provide valuable results. In addition to focusing on Witken's Field Dependence, instead of Pask's Holism/Serialism, it is, however, difficult to use the aforementioned methods to discover meaningful relationships between items or features that co-occur frequently in databases. For instance, it is not easy to identify which navigation tools are always used together to obtain the information from the WBLS. To this end, this study uses another data mining approach named association rule. Association rule was firstly proposed by Agrawal and Srikant (1994). It is mainly used to find out the meaningful relationships between items that co-occur frequently in databases. The purpose of the association rules is to investigate the meaningful relationships between items or features that occur together frequently. Therefore, it provides a useful way to predict the learners' navigation behavior based on their cognitive styles. In other words, this technique is helpful to reach the two-folds aims of this study.

3 Methodology

***WBLS*:** To identify learners' navigation behaviors, a WBLS is developed and the subject content is related to *Interaction design.* Multiple navigation tools were provided, including a main menu, keyword search, hierarchical map, and alphabetical index. The layout of the WBLS is shown in Fig.1.

Fig. 1. The WBLS

***Participants*:** For finding reliable evidence about the relationships between human factors and the use of navigation tools, 50 students from a University in Taiwan participated in our study. These participants were undergraduate or graduate students and they agree to take part in this study voluntarily. All participants have basic computer and Internet skills that are necessary to use the WBLS.

***Procedure*:** The procedure is composed of four steps: (1) Using Ford's study preference questionnaire (SPQ) to identify learners' cognitive styles (Ford, 1985); (2) Filling out learners' personal information; (3) Browsing the WBLS and finishing the required task simultaneously; and (4) Evaluating learners' performance. Each participant has a log file, which records his/her personal detail and their navigation behavior, including total time spent for interacting with the system and the frequencies of accessing each navigation tool. These data were used to analyze the relationships between learners' cognitive styles and their navigation behavior by using the association rules.

***Data Analysis*:** In this study, the association rule was used to conduct data analyses. More specifically, the association rule was employed to analyze the participants' demographic information, time spend interacting and the frequencies of using each navigation tool. Among a variety of techniques used for association rules, an *Apriori* algorithm is widely used to find all frequent item sets and generate the desired rules (Agrawal and Srikant, 1994) so it is adopted to reach our aims.

4 Results and Discussions

In this section, the results were discussed. Before doing so, the input data of the *Aprior* algorithm need to be defined firstly. As showed in Section 3, there are six features, including cognitive styles, time that the learners spend interacting with the WBLS, and the frequency of using each navigation tool. They are used as the input data in the *Aprior* algorithm, where nominal data type is required. Therefore, we need to translate these six features to the nominal data types. After preprocessing the input data, the *Aprior* algorithm is used to find out the association rules of the input data. Subsequently, four association rules are found and each association rule has the confidence above 0.8 (Table 1). As shown in Table 1, the association rules can be further divide into two major parts, i. e. holists' navigation behavior (Section 4.1) and serialists' navigation behavior (Section 4.2). The details of each part are shown below.

Table 1. The association rules

Rule ID	Confidence	Definition
R1	80%	*If holists prefer to use the index few times and the keyword search a few times, they may prefer to use the main menu many times.*
R2	90%	*If holists prefer to use the main menu many times and the map a few times, they may use the index few times.*
R3	92%	*If serialists prefer to use the main menu few times, they may prefer to use the keyword search many times.*
R4	89%	*If serialists prefer to use the map few times, they may use the index a few times.*

4.1 Holists' Navigation Behavior

According to Rule 1, holists, who use the keyword search a few times and never use the index, may prefer to use the main menu many times. Moreover, R2 indicates that holists, who tend to use the main menu many times and use the map a few times, may never use the index. In other words, holists may prefer to use the main menu but never prefer to use the index. It implies that the main menu, which allows learners to locate the information with "topic" based learning, is suitable for the holists. It is due to the fact that holists feel more comfortable with "topic" based learning (Ford et al., 2002). Conversely, the index, which provides single fragment item, does not meet the holists' needs. For further investigating their navigation behavior, the usage of each navigation tool is shown in Fig. 2. As shown in this figure, holists mainly use the main menu. It echoes R1 and R2, which indicate that holists prefer to focus on the particular navigation tools that allow learners to get a global picture in an early stage.

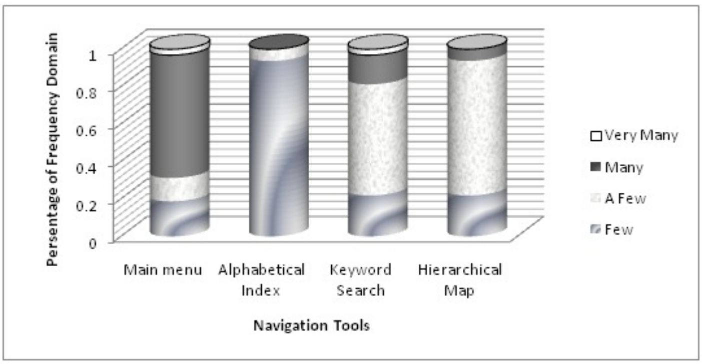


Fig. 2. The usage of each navigation tool (holists)

In addition, R1 and R2 also indicate that holists prefer to use the multiple navigation tools at the same time. Regarding the R1, holists prefer to use the main menu and keyword search at the same time. The difference between these two navigation tools lies within the fact that the former presents the overall picture of the subject content while the latter provides a single concept at a time. It implies that holists prefer to try different kinds of navigation tools at the same time. However, they just pay a little attention on the tools that are not suitable for them. Regarding R2, they prefer to use the main menu and map at the same time. These two tools share some similarities and differences. Regarding the similarities, such two tools provide an overall picture in an early stage. Regarding the differences, the former is suitable for novices to locate the information by using the rough idea and extend their reading by following the links while the latter is suitable for experts to look for information with specific concepts presented in the system. For further investigating this issue, we find that 82% of holists are novices. Thus, they may prefer to get an overall picture and follow the links presented in the link in an early stage (Pask, 1976). Such finding may also echo R1, which indicate that holists may prefer to use the keyword search a few times but never use the index. In short, holists, especially novices, prefer to use the rough idea to locate information, instead of using the specific concepts. Moreover, they may also prefer to get an overall picture in an early stage, instead of getting procedure details.

4.2 Serialists' Navigation Behavior

In addition to holists' navigation behavior, we also investigate the serialists' navigation behavior. According to R3 and R4, serialists display different navigation behavior when using different tools. More specifically, serialists prefer to focus on the keyword search and index, instead of using the main menu and map. The differences between the first two tools and the last two tools lie within the fact that the first two tools display the single concept at a time while the last two tools present the overall picture of the subject content. It is due to the fact that serialists prefer to build up their knowledge sequentially so they are not suitable for using the main menu and map which present the overall picture of the subject content (Pask, 1976). Conversely, the keyword search and index may suitable for them to get the single concept and understand procedure details at a time.

For further investigating their navigation behavior, the usage of each navigation tool is shown in Fig. 3. As shown in this figure, serialists mainly use the keyword search and index. It echoes R3 and R4, which indicate that serialists prefer to focus on the particular navigation tools that allow learners to build their knowledge sequentially. Moreover, we can also found that serialists use the keyword search more times than the index. The differences between the keyword search and index lies within the fact that the former allows learners to locate the information by using the rough idea while the latter provides specific concepts to help learners seek information easily. It implies that the index, which provides specific concepts, can help serialists easily focus on procedure details. Thus, they just need to use them a few times for seeking information. Conversely, the keyword search, which provides a rough idea, can help them easily find the concept by using their own ways. However, they may need to identify whether the information is useful or not frequently.

In addition, we also find that serialists also prefer to focus on a single navigation tool at a time when using the WBLS. It is also suitable for their characteristics that

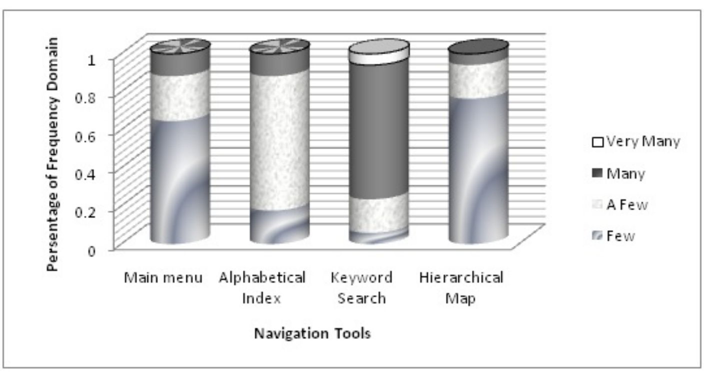


Fig. 3. The usage of each navigation tool (serialists)

they prefer to prefer to focus on a single item at a time. In short, serialists may prefer to not only focus on a single navigation tools at a time but also use the navigation tools that can help them build up their knowledge sequentially (i.e. the keyword search or index).

4.3 Relationships between Learners' Navigation Behavior and Their Learning Performance

Regarding learners' learning performance, we cannot find any rules relevant to learning performance. In other words, learning performance may be irrelevant to navigation behavior. It implies that the learners may prefer to use particular navigation tools according to their cognitive styles but these tools have nothing to do with their performance. A possible reason is that preference and performance are two different things. Preference represents how much a learner prefers a given tool or not while performance means the ability of learners actually solve problems (Topi & Lucas, 2005). Superior performance requires not only knowledge and skills but also how to use knowledge and skills effectively (Mavis, 2001). The findings suggest that there is no direct relationship between performance and navigation behaviors.

Based on the aforementioned four rules, a framework is proposed to summarize our findings (Fig. 4).

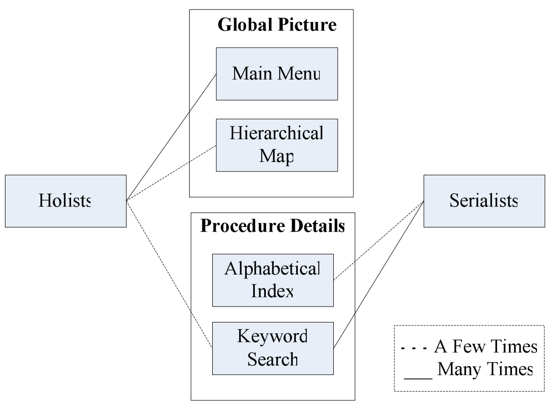


Fig. 4. The framework of this study

5 Conclusion

This study used the association rules to reach our aims. Regarding the first aim, the holists prefer to get an overall picture in an early stage and obtain information with multiple tools while serialists prefer to see a single concept and use only one tool at a time. Regarding the second aim, we found that there is no direct relationship between learning performance and preferences. In brief, learners' cognitive styles are important issues that should be considered when developing WBLSs. The findings have shown the importance of understanding the use of navigation tools in the WBLSs. However, this is a small-scaled study. Further work needs a larger sample to provide additional evidence to verify the findings presented in this study. In addition, this paper only presents the results from the *Apriori's* association rule so other data mining methods (e.g. clustering or classification) can also be used to investigate our aims.

Acknowledgements

This work is funded by National Science Council, ROC (NSC 98-2511-S-008 -012 -MY3 and NSC 99-2511-S-008 -003 -MY2).

Reference

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Databases, pp. 487–499 (1994)
2. Chen, S.Y., Liu, X.H.: Mining Students' Learning Patterns and Performance in Web-based Instruction: A Cognitive Style Approach. Interactive Learning Environments 1, 1–14 (2009)
3. Chen, S.Y., Liu, X.: An Integrated Approach for Modeling Learning Patterns of Students in Web-based instruction: A Cognitive Style Perspective. ACM Transactions on Computer-Human Interaction 15(1), Article 1 (2008)
4. Ford, N., Ford, R.: Toward a cognitive theory of information accessing: an empiric study. Information Processing & Management 29, 569–585 (1993)
5. Clewley, N., Chen, S.Y., Liu, X.: Cognitive Styles and Search Engine Preferences: Field Dependence/Independence vs. Holism/Serialism. Journal of Documentation 66(4), 585–603 (2010)
6. Ford, N.: Style and Strategies of Processing Information: Implications for Professional Education. Education for Information 3(1), 15–32 (1985)
7. Ford, N., Wilson, T.D., Foster, A., Ellis, D., Spink, A.: Information seeking and mediated searching. Part 4. Cognitive styles in information seeking. Journal of the American Society for Information Science and Technology 53(9), 728–735 (2002)
8. Han, J., Kamber, M.: Data mining, concepts and techniques: Cluster analysis. Morgan Kaufmann, San Francisco (2001)
9. Jonassen, D.H., Grabowski, B.: Individual differences and instruction. Allen and Bacon, New York (1993)
10. Lee, C., Cheng, Y.W., Rai, S., Depickere, A.: What affect student cognitive style in the development of hypermedia learning system? Comput. Educ. 45(1), 1–19 (2005)
11. Li, H., Chen, H.: Mining non-derivable frequent itemsets over data stream. Data & Knowledge Engineering 68(5), 481–498 (2009)
12. Mavis, B.: Self-Efficacy and OSCE Performance Among Second Year Medical Students. Advances in Health Sciences Education 6(2), 93–102 (2001)
13. Minetou, C.G., Chen, S.Y., Liu, X.H.: Investigation of the Use of Navigation Tools in Web-Based Learning: A Data Mining Approach. International Journal of Human-Computer Interaction 24(1), 48–67 (2008)
14. Pask, G.: Styles and strategies of learning. British Journal of Educational Psychology 46, 128–148 (1976)
15. Topi, H., Lucas, W.: Mix and match: combining terms and operators for successful Web searches. Information Processing & Management 41(4), 801–817 (2005)
16. Wang, C., Wang, D.Z., Lin, J.L.: ADAM: An adaptive multimedia content description mechanism and its application in web-based learning. Expert Systems with Applications 37(12), 8639–8649 (2010)

The Design of Adaptive Error Feedback Music Ear-Training System with Image Cues

Yu Ting Hwang[1] and Chi Nung Chu[2]

[1] Shih Chien University, Department of Music, No.70 Ta-Chih Street, Chung-Shan District,Taipei, Taiwan, R.O.C.
[2] 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 AEFMES (Adaptive Error Feedback Music Ear-Training System). It is an assessment analysis engine designed to provide immediate diagnostic feedback on the melodic line assessment with pitch recognition, interval recognition and rhythm recognition. The AEFMES with image cues could move learners beyond basic drill exercises to a competence that is tailored to the content of individual needs in the ear-training process. Many misconceptions of students can be cleared up through the combination of practice and immediate adaptive error feedback with image cues.

Keywords: Ear-Training**,** Pitch Recognition, Interval Recognition, Rhythm Recognition.

1 Introduction

Music ear-training includes identifying pitch, intervals, chords and rhythm which are essential elements to successful music world. The inherent abstract complexity of extraction of identification in musical variation is hard for learners to comprehend and learn. The move to use technology to support learning has become an emerging development in the recent music pedagogy [4, 5]. Many learners in traditional learning environment have limited immediate learning feedback which allows them to review, reinforce, and develop such aural skills. As the difficulty of the music texture increases, so does their frustration. Therefore providing mistake analyses in music ear-training learning is needed to consolidate their own aural skills [2, 6].

Musical auditory processing with image cues possesses educational implications. Studies showed the effects of verbal labels on music recognition could facilitate memory recall of musical passages [1, 3]. This paper tried to use technology to practice using assessment analysis engine with image cues as individual's error feedbacks. Learners could facilitate their ear-training through self- drill tasks.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 35–38, 2011.

2 Adaptive Error Feedback Music Ear-Training System

The Adaptive Error Feedback Music Ear-Training System (AEFMES) in this study, integrating the Microsoft Agent as a verbal tutor, is composed of assessment analysis engine (Fig. 1) and piano drill engine (Fig.2). The assessment analysis engine is designed to provide immediate diagnostic feedback on the melodic line assessment with pitch recognition, interval recognition and rhythm recognition. For pitch recognition or interval recognition practice, each item in the online assessment responds to specified pre-analyzed error type respectively. And each error type in pitch recognition and interval recognition is further classified into subtypes. Each feedback of error type associates with image cue in different color, and each feedback of error subtype associates with image cue in different shape.

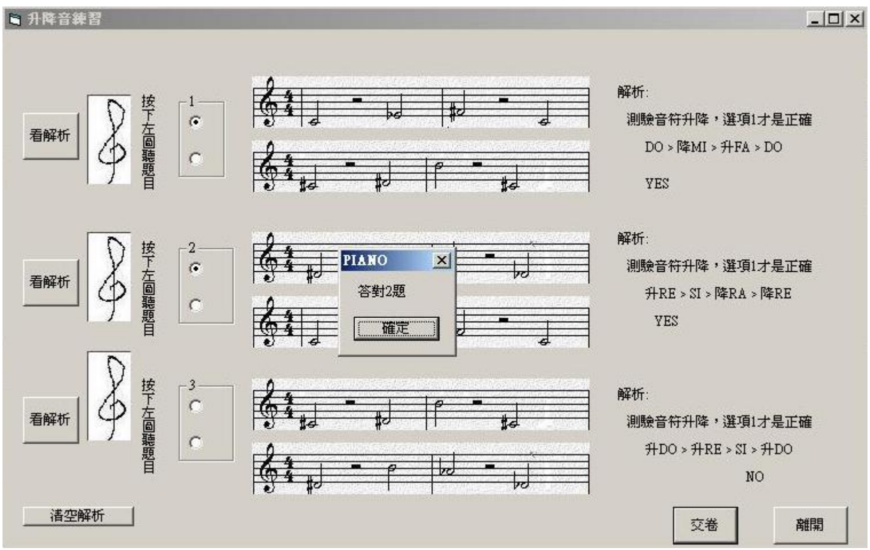


Fig. 1. Assessment analysis engine

Learners could retrieve personally error feedbacks that they made during the assessment process. All the practice items together with corresponding information are saved into XML file. The assessment analysis engine could thus analyze and search the XML tree to provide individual feedbacks. And the piano drill engine is designed to provide immediate practicing feedback on the melodic line assessment with pitch recognition, interval recognition and rhythm recognition.

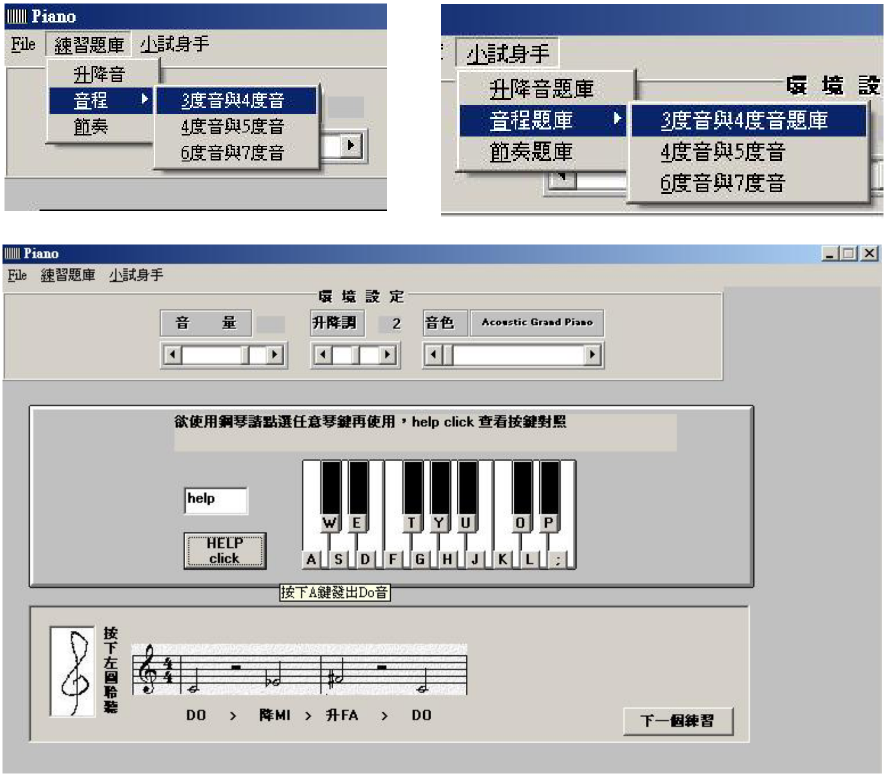


Fig. 2. Piano drill engine

3 Evaluation of AEFMES

The subjects in the study consisted of 23 university students who do not have a strong music background. The experiment involved the use of AEFMES integrated into the music interval instruction program. The assessment analysis engine could distinguish total fourteen major error types for each pitch recognition and interval recognition respectively that were verified with expert validity. Each error type associates with different color as the image cue background for the assessment feedback, including red, orange, yellow, green, blue, white and purple. And each subtype error which could be further analyzed by the assessment analysis engine associates with different shape, including circle, triangle, square and diamond in corresponding background color. After every assessment analysis, there will pop up a verbal agent with tutoring information and image cue.

Learners were given three different musical intervals to recognize (Minor 2nd, Major 3rd, and Perfect 4th). While listening to the intervals they could request piano drill engine to practice what that interval sounds like. According the personal error feedback from the assessment analysis engine, learner could then focus interval recognition on verbal tutoring information with the image cue. And the specific image cue could be triggered for the corresponding error type of practicing with piano drill engine. A post-test was administered with all the data collected to a server database.

The initial analysis showed a significant increase from the pre to post-test scores ($F=7.75, P<.05$), indicating that the AEFMES with image cues was effectively teaching the given material. And Pearson correlations were conducted between the post-treatment scores. The results showed no significant relationships existed. The image cue does not exist in the relationship between an auditory occurrence and the visual signal of that occurrence, but the learner had been conditioned to create something meaningful to themselves that could be used as a visual cue in the treatment.

4 Conclusions

The AEFMES with image cues could move learners beyond basic drill exercises to a competence that is tailored to the content of individual needs in the ear-training process. Music ear-training development occurs when learners interact with AEFMES in a continuous drill. Many misconceptions can then be cleared up through the combination of practice and immediate adaptive error feedback with image cues.

References

1. Clark, J.M., Paivio, A.: Dual coding theory and education. Educational Psychology Review 3(3), 149–210 (1991)
2. DeCarbo, N.J.: The effects of conducting experience and programmed materials on error-detection scores of college conducting students. Journal of Research in Music Education 30, 187–200 (1982)
3. Henik, G.: Music and Self-Generated Images: Applying Dual Coding Theory, Generative Theory of Reading Comprehension, and Cognitive Load Theory to Web-Based Music Instruction. In: Lassner, D., McNaught, C. (eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, pp. 1129–1136. AACE, Chesapeake (2003)
4. Hoffstetter: GUIDO: an interactive computer-based system for improvement of instruction and research in ear-training. Journal of Computer-Based Instruction 1(4), 100–106 (1975)
5. LeBlanc, A., Jin, Y.C., Obert, M., Siivola, C.: Effect of audience on music performance anxiety. Journal of Research in Music Education 45(3), 480–496 (1997)
6. Ramsey, D.S.: Programmed instruction using band literature to teach pitch and rhythm error detection to music education students. Journal of Research in Music Education 27, 149–162 (1979)

Fuzzy Linguistic Modelling Cognitive / Learning Styles for Adaptation through Multi-level Granulation

Ilham N. Huseyinov

Computer Engineering Department, European University of Lefke,
Gemikonagi-Lefke, Mersin 10, Turkey
ihuseynov@yahoo.com

Abstract. In this study, based on fuzzy linguistic modelling and fuzzy multi level granulation an adaptation strategy to cognitive/learning styles is presented. Fuzzy if-then rules are utilized to adaptively map cognitive/learning styles of users to their information navigation and presentation preferences through natural language expressions. The important implications of this approach are that, first, uncertain and vague information is handled; second, a mechanism for approximate adaptation at a variety of granulation levels is provided; third, a qualitative linguistic model of adaptation is presented. The proposed approach is close to human reasoning and thereby lowers the cost of solution, and facilitates the design of human computer interaction systems with high level intelligence capability.

Keywords: adaptive hypermedia, fuzzy granulation, cognitive/learning styles, navigation/presentation preferences, linguistic variables.

1 Introduction

The growing amount of information on the WEB and the heterogeneous characteristics of Web users have lead to a considerable attention to web-based adaptive hypermedia systems (WAHS) by the research community. The power of hypermedia of web technology is in its capability to support non-linear navigation in hyperspace and multimedia presentation of the web content. WAHS offers an alternative to the traditional "one-size-fits-all" hypermedia and Web systems by adapting to the goals, interests, and knowledge of individual users represented in the individual user models [1]. WAHS aims to minimize cognitive overload faced by users, to alleviate the disorientation problem of users, to enhance the usability and the utility of the system by applying intelligent information adaptation (personalization) techniques for user/system interactions that take into account individual differences of users [2]. Adaptation involves two key activities: (i) a user modelling activity to develop a user model and (ii) an adaptation activity that leverages a 'rich' user-model to personalize the information content, the information presentation style and the navigation path of the system to the user [3]. One of the ways to enhance the efficiency of WAHS is to build accurate user models. It can be achieved by taking into account human factors (or individual differences) that have significant effects on

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 39–47, 2011.

human computer interaction and on the learning process [4, 5]. Research into individual differences suggests cognitive/learning styles (CLS) have significant effects on student learning in hypermedia systems [5- 8].

Information imperfection, that is, information used in one or more respects is imprecise, uncertain, incomplete, unreliable, vague or partially true, is an inherent characteristics of WAHS [9]. Imprecision of WAHS is rooted in imprecision of its input information. Sources of input information to build a user model are either the explicit information provided by subjective judgments of users/experts or the implicit information inferred by monitoring and measurement of a user behaviour, or combination of both. This imprecision is passed on user model and then on adaptation strategy that is guided by heuristics, hypotheses, or approximate decisions.

This paper deals with uncertain and vague information in WAHS and proposes fuzzy logic approach to handle it. Fuzzy granulation method is proposed to create a fuzzy inference system (FIS) to adaptively map cognitive/learning styles of users to their information navigation/presentation preferences (NPL) through natural language expressions. The important implications of this approach are that, first, uncertain and vague information used is handled; second, a mechanism for approximate adaptation at a variety of granulation levels is provided; third, a qualitative linguistic model of adaptation is presented. The proposed approach is close to human reasoning and thereby lowers the cost of solution, and facilitates the design of human computer interaction systems with high level intelligence capability.

The paper is organized as follows. The description of cognitive and learning styles is given in Section 2. Navigation and presentation preferences of users are presented in Section 3. The adaptation process, the fuzzy granulation of input and output linguistic variables, and an inference mechanism are presented in Section 4. Section 5 is devoted to related works. Finally, in Section 6 we present results and conclusion.

2 Cognitive/Learning Styles

Cognitive styles (CS) deal with the form of cognitive activity (thinking, perceiving, remembering), not its content. Learning styles (LS), on the other hand, is seen as a broader construct, which includes cognitive along with affective and physiological styles. A key factor in determining cognitive styles with respect to learning is the field dependency factor. Filled dependence refers to an individual's ability to perceive a local field as discrete form of its surrounding field. It is a single bi-polar dimension ranging from Field dependent (FD) individuals at one extreme to Field independent (FI) individuals at the other [10]. Some cognitive abilities are highly coupled to perceptual abilities because they are sensitive to acquisition modalities (view, hearing, touch, taste, smell) [11]. Characteristics of users in respect to CL are described in [7] and can be modelled by a hierarchy type tree structure given in Fig. 1.

The dimension *visual/verbal* of the model developed by Felder- Silverman is used to reflect the learning modalities. However, CLS is a disputable concept that is not fully accepted by the whole community.

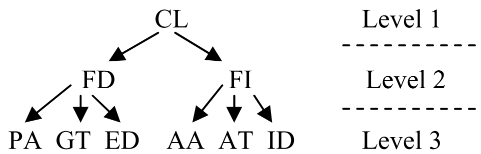


Fig. 1. Hierarchical type tree structure of CLS

where FD – field dependent, PA – passive approach, GT – global tendency, ED – externally directed, FI – field independent, AA – active approach, AT – analytical tendency, ID – internally directed.

In most of the systems CLS is assessed through psychological questionnaires and psychometric tests or in the form of self-report. This kind of measures of CLS is based on subjective judgment users make about themselves. Furthermore, not all characteristics they measure are stable and invariable across different subject domains. It is often the case when the mixed result for the same person is obtained, that is a user may have preference for one particular style, preference for more than one style and different levels of preferences for the different styles [12, 13]. For example, a learner may be attributed to the visual style at the high level, but also to verbal style up to a certain extent, at the medium level. For that reason, CLS characteristics of users are intrinsically imprecise and consist of overlapping classes of styles one can not to draw a line between them. Fuzzy logic and granulation methods are proposed in Section 4 to handle the fuzziness of CLS.

3 Navigation and Presentation Preferences

Navigation preferences of users in respect to cognitive styles characteristics described in Fig. 1 can be presented in the form [7]:

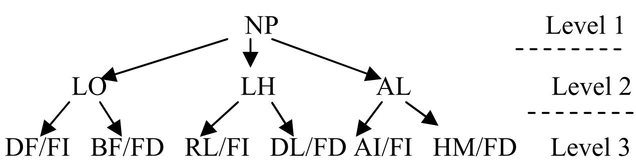


Fig. 2. Navigation preferences

where NP – navigation preference, LO – link ordering, LH – link hiding, AL – adaptive layout, DF – depth firth path, FI – field independent, BF – breadth-firth path, FD – field dependent, RL – rich links, DL – disabled links, AI – alphabetical index, HM – hierarchical map.

Presentation preferences are modes of delivering the content using a variety of multimedia techniques such as text, graphics, image, audio, video and etc.

4 Adaptation to Cognitive/Learning Styles

Two types of adaptation are distinguished in WAHS, namely adaptive presentation support and adaptive navigation support. The aim of adaptive navigation is to support users to find their learning paths in hyperspace by editing links. For example, a link can be added, removed, or edited to change its format and presentation. The adaptive presentation supports users in selecting the content of the current node of course structure and the content presentation style or mode [3]. In this study, the basic source of the adaptation information is CLS that is to be adapted to navigation and presentation preferences of users. We interpret the process of adaptation as an inference mechanism that maps a set of input variables to a set of output variables. FIS provides the formalism that maps one family of fuzzy sets to another family of fuzzy sets. This formalism serves as a means of precision of imprecise information through graduated (or fuzzy) granulation [14, 15].

4.1 Fuzzy Multi Level Granulation of Input Linguistic Variables

Let us introduce linguistic variables ***CS, FD, PA, GT, ED, FI, AA, AT, ID,*** associated with the concepts CS, FD, PA, GT, ED, FI, AA, AT, ID, respectively, which are described in Fig. 1. Membership functions of fuzzy sets of these linguistic variables are defined by applying multi level granulation technique. The first level granulation applied to the root ***CS*** of the tree yields:

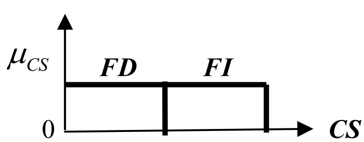


Fig. 3. Crisp sets for linguistic variable ***CS***

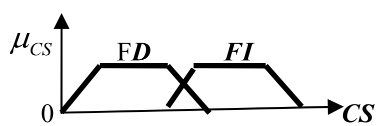


Fig. 4. Granulation of ***CS***

The second level of granulation is applied to the nodes of the tree that are in the second level, for example, for ***FD*** we have:

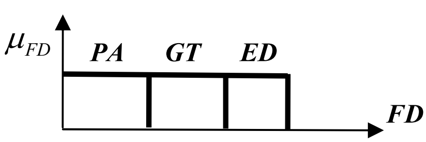


Fig. 5. Crisp sets for ***FD***

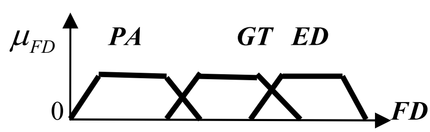


Fig. 6. Granulation of ***FD***

Finally, the granulation of the nodes from the third level through linguistic qualifiers ***poor, good,*** and ***excellent*** is shown in Fig. 7.

Next, we can use *hedges* for the next granularity level. Hedges are linguistic modifiers operated on membership functions. They are expressed by adjectives and/or adverbs such as *very, somewhat, slightly, more or less, quite, extremely, fairly, below* and etc. For example, hedge *very* applied to qualifier *poor* modifies its membership function as follows: $\mu_{verypoor} = (\mu_{poor})^2$.

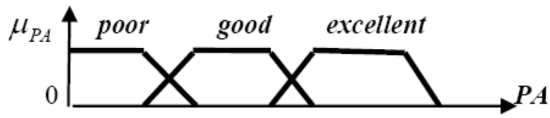


Fig. 7. Granulation of linguistic variable ***PA***

The *visual/verbal* styles modelled by linguistic variables ***VI*** and ***VB***, respectively, can also be granulated in the similar way.

4.2 Fuzzy Multi Level Granulation of Output Linguistic Variables

Based on Fig. 2 we introduce output linguistic variables ***NP, LO***, ***LH, AL, DF, BF, RL, DL, AI,*** and ***HM*** associated with the concepts NP, LO, LH, AL, DF, BF, RL, DL, AI, and HM, respectively. Applying multi level granulation method, similar to the one described above, we can form output membership functions for output linguistic variables ***LO***, ***LH,*** and ***AL*** in the form shown in Fig. 8, where for the linguistic variable ***LO*** - ***A*** stands for ***DF***, ***B*** stands for ***BF;*** for the linguistic variable ***LH* - *A*** stands for ***RL***, ***B*** stands for ***DL***; and finally, for the linguistic variable ***AL*** - ***A*** stands for ***AI***, B stands for ***HM.***

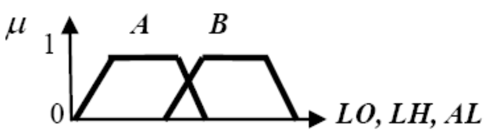


Fig. 8. Granulation of linguistic variables ***LO***, ***LH,*** and ***AL***

Finally, ***we*** create one more output linguistic variable ***MMM*** associated with the concept *presentation preferences* and relate it with the input linguistic variables ***VI/VB***. The multimedia mode of presentation includes text, image, audio, video, graphics, games, animation and any combination of these elements. A granulation of linguistic variable ***MMM*** can be shown as follows:

Fig. 9. Granulation of the linguistic variable MMM

Again, we can use linguistic qualifiers and modifiers, as we did with input linguistic variables, for the next granulation levels. The linguistic qualifiers *low, medium, high* used for output linguistic variables are shown in Table 1.

4.3 Inference

Fuzzy inference is a method that interprets the values in the input vector and, based on some set of if -then rules, assigns values to the output vector. Fuzzy if - then rules capture the expert knowledge in the form of fuzzy predicates that establishes relations between input and output linguistic variables. The fuzzy predicates are associated with linguistic terms, and the proposed model is, in fact, a qualitative description of the system using rules like: IF input linguistic variable ***CLS*** is *poor* THEN output linguistic variable ***NPL*** is *high*. Such models are often called linguistic models [16, 17].

Table 1. FIS for adaptation cognitive/learning styles to navigation/presentation preferences with a variety of granulation levels

Input linguistic variables			terms	hedges	rules	hedges	terms	Output linguistic variables		
Granular levels								Granular levels		
1	2	3						3	2	1
			poor	***Very***	r1	***very***	***low***	***text***		
				slightly	r2	***slightly***		***graphics***		
	VI					***somewhat***		***image***	***MMM***	
					-					
				somewhat				***audio***		
										NPL
CLS					-	***more or less***				
								video		
	VB		***good***	***more or less***			***medium***			
					-					
		PA						***DF***		
						quite				
	FD	***GT***			-			***BF***	***LO***	
				quite						
		ED						***RL***		
					rn	***extremely***				
		AA						***DL***	***LH***	
				extremely						
	FI	***AT***	***excel***				***high***	***AI***		
		ID	***lent***					***HM***	***AL***	

More formally, let ***CLS*** and ***NPL*** are linguistic variables defined by fuzzy sets on the universes of discourse that contain granular values described in 4.1 and 4.2. Denote membership functions of linguistic variables ***CLS*** and ***NPL*** by μ_{CLS} and μ_{NPL}, respectively. Then the adaptation process can be characterized by a mapping $f : \mu_{CLS} \rightarrow \mu_{NPM}$. Granulation of function f is a fuzzy graph that is described as a collection of if-then rules. A fuzzy if-then rule can be defined as a binary fuzzy relation R considered as a fuzzy set with membership function: $\mu_R = f(\mu_{CLS}, \mu_{NPL})$. Using the compositional rule of inference, we can formulate the inference procedure in fuzzy reasoning in the form: $\boldsymbol{NPL} = \boldsymbol{CLS} \circ R$, where the sign $\circ$ denotes a fuzzy composition operator, consisting of a t-norm operator, followed by a t-conorm operator. The FIS for the adaptation of CLS to NPL is shown in Table 1. Some examples of rules at a variety of granulation level are presented below:

Level 1: IF *CLS* is *FD,* THEN *NPL* is *DL*; IF *CLS* is *AA* , THEN *NPL* is *DF;* IF *CLS* is *poor,* THEN *NPL* is *high*; IF *CLS* is *very poor,* THEN *NPL* is *extremely high.*

Level 2: IF *FD* is *good,* THEN *DL* is *medium*; IF *FI* is *quite good,* THEN *AI* is *somewhat medium;* IF *VI* is *excellent,* THEN *MMM* is *video;* IF *VI* is *good* and *VB* is *good,* THEN *MMM* is *video* and *audio*.

Level 3: IF *PA* is *good* and *AA* is *good,* THEN *DF* is *medium* and *BF* is *medium;* IF *AA* is *very good,* THEN *AI* is *quite high.*

The number of rules increases exponentially with the number of inputs, but some rules may rarely or never occur. So, fewer rules may be predefined by experts in the domain. Moreover, one can switch between granules at different levels. It can be done using a meta rule. For example, if *X* is a set of granules *FD, FI, VI* and *VB*, then the following two step rule can be written: IF *CLS* is *X,* THEN *select the rule associated with the value of X.* Here the first rule serves as a meta rule and is used to control the order of firing rules from Level 2.

5 Related Work

Fuzzy modelling has previously applied with success to a variety of problems of WAHS. FIS based on a rule-based representation of user profiles is given in [18] for web recommendation system. In the educational AHS, the work [19] employs fuzzy logic for modelling user knowledge of domain. The work [20] proposed a model named MAZE to capture imprecise behaviours of users. The neuro-fuzzy approach successfully applied to the student diagnosis in [21]. The work [22] proposes an algorithm of adaptation to psychological factors using multi utility attribute theory and fuzzy set theoretic similarity measure. A survey of application of soft computing methods to modelling user behaviour is presented in [23]. The neuro-fuzzy based web recommendation system is examined in [24].

6 Results and Conclusion

The proposed linguistic model can be simulated using Fuzzy Logic Toolbox in Matlab. The aim is to observe the behaviour of the proposed model and tune its parameters such as: input and output linguistic terms and their membership function shapes; relevance and weights of rules; the type of inference mechanism – Mamdani type max-min composition or Takagi-Sugeno type linear bounded-sum; the type of defuzzification (centroid, middle of maximum, largest of maximum, and smallest of minimum). The Mamdani type FIS expects the membership functions of output linguistic variables to be fuzzy sets and requires the defuzzification step, while Takagi-Sugeno type FIS expects output membership functions to be singletons. Until now, preliminary simulation results have shown that FIS based on Mamdani max-min is preferred at a lower level of granulation, whiles Sugeno-type FIS is more appropriate at a higher level of granulation. The simulation results are in line with expectations of domain experts. As a future work, a prototype of the model shall be developed to validate the proposed linguistic model within the efficiency of user/system interactions in terms of the usability and the utility of the system. This paper has proposed a fuzzy linguistic model for adaptation to CLS based on knowledge acquisition through multi level granulation method. The implication of this approach for designing adaptive user interfaces is introducing more natural language expressions. The qualitative description of the adaptation process is close to human reasoning and thus facilitates the design of the adaptive interface, increases the interpretability, usability of the system and utility of information, saves memory space and cost expenses by computation with words instead of numbers.

References

1. Brusilovsky, P.: Adaptive Hypermedia. User Modelling and User- Adapted Interaction 11(1/2), 87–110 (2001)
2. Mobasher, B. Anand S. S.: Intelligent techniques for Web personalization, `http://www.inf.unibz.it/~ricci/ATIS/papers/itwp-v5.pdf`
3. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modelling and User- Adapted Interaction 6(2), 87–129 (1996)
4. Tsianos, N., Germanakos, P., Lekkas, Z., Mourlas, C., Samaras, G.: An assessment of human factors in adaptive hypermedia environments, pp. 1–34. IGI Global (2009)
5. Triantafillou, E., Pomportsis, A., Demetriadis, S.: The Design and the Formative Evaluation of an adaptive Educational System based on Cognitive Styles. Computers & Education 41, 87–103 (2003)
6. Chen, S.: A cognitive model for non-linear learning in hypermedia programmes. British Journal of Educational Technology 33(4) (2002)
7. Mitchell, T., Chen, S.Y., Macredie, R.: Cognitive Styles and Adaptive Web-based Learning, `http://bura.brunel.ac.uk/handle/2438/388`
8. Tsianos, N., Germanakos, P., Lekkas, Z., Mourlas, C.: Individual differences in Adaptive educational hypermedia: the effect of cognitive style and visual working memory, pp. 147–163. IGI Global (2009)
9. Garcia, E., Sicilia, M.A.: Information imperfection as inherent characteristics of adaptive hypermedia: imprecise models of users and interactions. In: Chen, S., Magoulas, G. (eds.) Adaptable and Adaptive Hypermedia Systems, pp. 150–167. IBM Press (2010)
10. Witkin, H.A., Moore, C.A., Goodenough, D.R., Cox, P.W.: Field-dependent and Field independent Cognitive Styles and Their Educational Implications. Review of Educational Research 47, 164 (1977)

11. Tarin-Bernard, F., Habieb-Mammar, H.: Modelling Elementary Cognitive Abilities for Adaptive Hypermedia Presentation. User Modelling and User-Adapted Interaction 15, 459–495 (2005)
12. Coffield, F., Moseley, D., Hall, E. and Ecclestone, K.: Learning Styles and Pedagogy in Post-16 Learning: a Systematic and Critical Review. Learning and Skills research centre, http://www.lsda.org.uk/files/pdf/1543.pdf
13. Stash, N.: Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System. Eindhoven: Technische Universiteit Eindhoven, Proefschrift (2007), http://alexandria.tue.nl/extra2/200710975.pdf
14. Zadeh, L.A.: Is there a need for fuzzy logic? Information Sciences 178, 2751–2779 (2008)
15. Zadeh, L.: The concept of a linguistic variable and its applications to approximate reasoning. Part 1, Information sciences 8, 199-249 (1975), Part 2, Information Sciences, 8 301– 357(1975), Part 3, Information Sciences, 9, pp. 43-80 (1975)
16. Zimmerman: Fuzzy set theory and its applications. Kluwer Academic Publishers, Dordrecht (2001)
17. Siler, W., Buckley, J.J.: Fuzzy expert systems and fuzzy reasoning. John Wiley & Sons, Chichester (2005)
18. Nasraoui, O., Petenes, C.: Combining web usage mining and fuzzy inference for web personalization. In: Proceedings of WEBKDD 2003: Web Mining as Premise to Effective Web Applications, pp. 37–46 (2003)
19. Kavcic, A.: Fuzzy User Modelling for Adaptation in Educational Hypermedia. IEEE Transaction on Systems, Man, and Cybernetics – part C: Applications and Reviews 34(4) (2004)
20. Garcia, E., Sicilia, M.A.: Information imperfection as inherent characteristics of adaptive hypermedia: imprecise models of users and interactions. In: Chen, S., Magoulas, G. (eds.) Adaptable and Adaptive Hypermedia Systems, pp. 150–167. IBM Press (2010)
21. Stathacopoulou, R., Grigoriadou, M., Samarakou, M., Mitropoulos, D.: Monitoring student's action and using teacher's expertise in implementing and evaluating the neural networked-based fuzzy diagnostic model. Expert Systems with Applications 32, 955–975 (2007)
22. Huseyinov, I.N., Cellatoglu, A.: Adaptation based on psychological factors using fuzzy logic. In: Proceedings of Int'l Conf. e-Learning, e-Bus., EIS, and e-Gov., EEE 2010, pp. 297–303 (2010)
23. Frias-Martinez, E., Magoulas, G., Chen, S., Macredie, R.: Modelling human behaviour in user adaptive systems: Recent advances using soft computing techniques. Expert Systems with Application 29, 320–329 (2005)
24. Castellano, G., Fanelli, A., Torsello, M.: Dynamic link suggestion by a neuro-fuzzy web recommendation system. In: IADIS International Conference WWW/Internet, pp. 219–225 (2006)

Method for Cultivating the "Inquiry-Mindset" Using the Information Access-Based Belief Bias Parameter

Kyoko Ito[1,2], Yuki Ito[2], and Shogo Nishida[2]

[1] Center for the Study of Communication-Design, Osaka University
[2] Graduate School of Engineering Science, Osaka University
{ito,yuuki,nishida}@nishilab.sys.es.osaka-u.ac.jp

Abstract. In today's world, it is important to have an "inquiry-mindset", i.e. the disposition to access information in a critical manner. There are several previous approaches that use discussion for cultivating an "inquiry-mindset". However, it is not easy to cultivate an inquiry-mindset only by discussion. This study proposes a method for cultivating an inquiry-mindset that focuses on opening the learner's eyes to the possibility of belief bias, that is the tendency to gather only agreeable information. This is a method for the quantification of belief bias based on three factors: accessing information, deciding the degree of importance, and viewpoint. Based on the evaluation experiment in which 38 students participated, the authors observed that accessing information and viewpoint are especially important for cultivating an inquiry-mindset. This result confirms the efficacy of the model and argues in favor of its combined use with other cultivating methods.

Keywords: Inquiry-mindset, Belief bias, Information access, Critical thinking disposition, Learning support.

1 Introduction

With the growing influence of science and technology on society, complex social problems appear, that cannot be solved only by science, and people are often required to reach social agreement [1]. It is necessary to critically gather multiphase information and to acquire a wide range of nonbiased information in order to achieve the appropriate mutual agreement. People need to acquire information critically through an attitude which is called critical thinking (CT) attitude [2]. The CT attitude can be further divided into four domains: awareness to logical thinking, inquiry-mindset, objectiveness and evidence based judgment [3]. The inquiry-mindset refers to the attitude of trying to acquire new information and different ideas, and is the most necessary factor for acquiring a wide range of nonbiased information. With the improvement of inquiry-mindset, the quality of social mutual agreement is also expected to improve.

There are some methods for cultivating the CT attitude using information technology as shown in related works [4] [5]. However, these studies have not yielded

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 48–57, 2011.

the expected results in the cultivation of an inquiry-mindset, failing to address issues such as internalizing information indiscriminately or ignoring information that contradicts one's standpoint. This study aims at proposing a method for cultivating the inquiry-mindset based on the advantages of both a discussion framework and information technologies. In addition, the factors that influence cultivating the inquiry-mindset are also analyzed by actually applying the proposed method to university students, and the further challenges for cultivating the inquiry-mindset are considered.

2 Method for Cultivating the Inquiry-Mindset by Stimulating Awareness to Belief Bias

2.1 Model for Cultivating the Inquiry-Mindset Using the Belief Bias Parameter

The inquiry-mindset has a great influence on people's acquisition of information and, therefore, for cultivating the inquiry-mindset two sides must be taken into account: the side where information is acquired (learner) and the side where the information is conveyed (information media). It is known that these two sides have different types of bias. First, there is the bias on the learner side (where information is acquired), in other words, the belief bias. Belief bias means the tendency to collect information and judge the right or wrong of things based on the agreement with one's own belief, not on the validity of logic [6]. Therefore, it is possible to suppose that there is a deep relationship between the inquiry-mindset and belief bias, and the authors assume that this can be the clue to cultivating an inquiry-mindset. On the other hand, there exists another type of bias, on the side that imparts information: media bias [7] [8]. Since the inquiry-mindset is the attitude of an individual towards information acquisition, it is more important to show the learner's own tendency to gather information, rather than only the results after removing the media bias. Therefore, it would be more effective to focus on the belief bias on the learners' side, and propose a mechanism that informs him/her of his/her own belief bias in order to cultivate an inquiry-mindset. Belief is variously defined according to each research field. In the field that focuses on the learner's understanding of utterances, it is defined as the matters or things that the learner considers true and valid [9]. Thus belief bias might become a factor that obstructs the access to relevant information. However, the belief bias is something generally possessed by everyone. This study refers neither the good/bad of the bias, nor does it deny or support bias avoidance. Here, information refers to things and matters including someone's intentions, hypotheses, and sense of values. Information access refers to the process by which someone tries to acquire information, and the awareness (or lack thereof) that information gathering is influenced by one's own viewpoint.

Based on the above, the authors assume that "Urging awareness to belief bias might lead to cultivating the inquiry-mindset". It is necessary to show that information access can be biased, in order to provide the learners with awareness to their belief bias. The determinism model, the probability theory, and the set-theoretical relation of Inami [10] can function as a theory connecting the elements concerning information and a person's sense of values. In this model, it is shown that the tendencies of

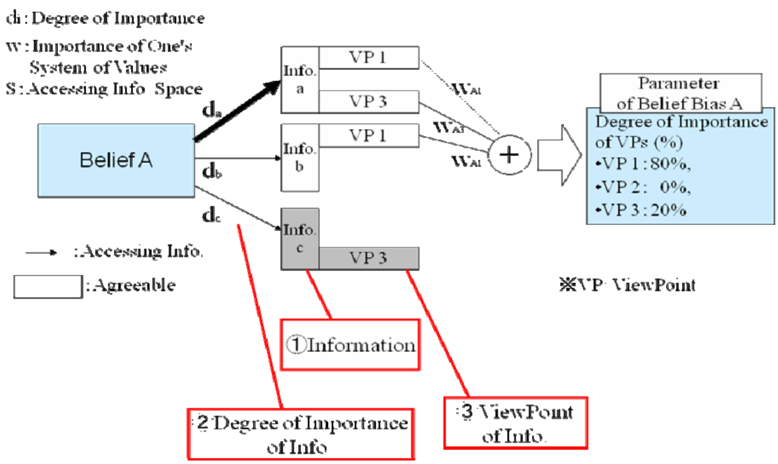


Fig. 1. A Model for the Quantification of Belief Bias

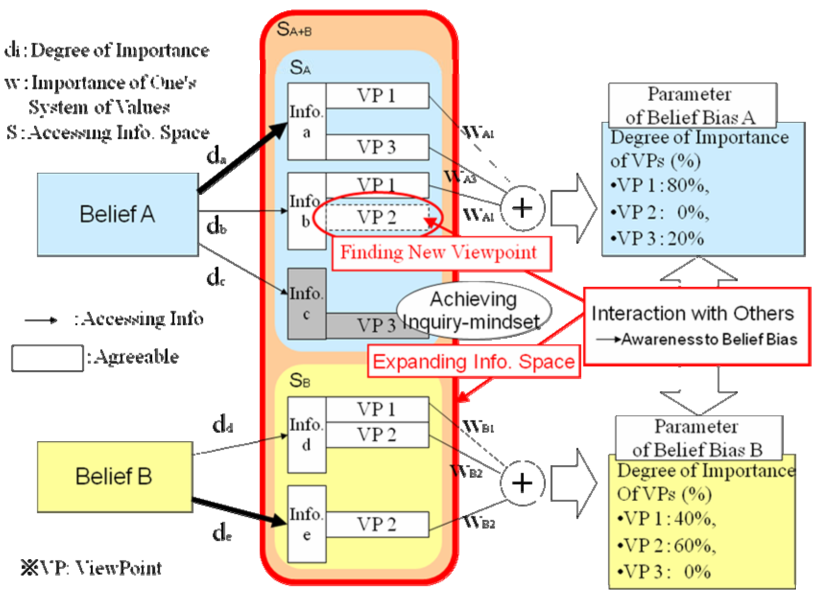


Fig. 2. A Model for Cultivating Inquiry-mind based on Belief Bias Parameter

attaching importance to information according to each learner's viewpoint can be quantified using the linear harmony of the values extracted from the weight and the viewpoint regarding a certain piece of information, on the one hand, and the way a person attaches importance to the topic, based on his/her sense of values, on the other. In addition, it is said that the information access bias appears not only for information acquisition but also as far as the priority level of information and the interpretation of information are concerned. Our challenge is to extract the information access bias by using three elements; information (itself), the degree of importance of the information, and the viewpoints from which the information are acquired.

In this study, the authors attempt to quantify the belief bias in connection with the influence of the information access bias and the role played by one's sense of values. The belief bias is quantified from the above-mentioned consideration points, aiming at cultivating an inquiry-mindset; also, the way it is used in the interaction with others is modeled. The belief bias quantification model is shown in Figure 1.

First, the bias about information ("Information") accessed by belief ("Belief") of learner A is expressed. Here, the arrows show that learner A is accessing information a, b, and c ("Info. a, b, c") according to the belief ("Belief A"). When the learner is supporting a certain belief (in the case of information a, b), this fact is shown by the difference in the color of the box (white color). The person's tendency is also expressed in the importance degree ("Degree of Importance") of the accessed information, where the importance degree is shown by the thickness of the arrow. For instance, "Info. a" is shown as having a higher degree of importance than "Info. b" by making the arrow to "Info. a" thicker. Moreover, "Viewpoint (ViewPoint)" shows the person's view about the information, that is, their interpretation of the accessed information. For instance, the fact that "Info. a" is interpreted from viewpoints 1 and 3 (VP1, VP3) is shown at the right side of the piece of information. For m number of pieces of information, the importance degree vector of information is **d** ($d_i = [d_{i1}, d_{i2}, ...d_{in}]$ (n: the number of viewpoints)), which is calculated by using the importance degree of the i-th information. The viewpoint vector is **v** ($v_i = [v_{i1}, v_{i2}, ...v_{in}](v_{ij} \in 0,1)$), which shows whether the information has a viewpoint or not. Since the sense of values from one individual to the other differs, the vector for the degree of importance of information is corrected linearly by the weight of each viewpoint determined by the individual's sense of values. Here, the weight of the viewpoint is **w** ($w = [w_1, w_2, ..., w_n]$). The belief bias parameter is **b** ($b = [b_1, b_2, ...b_n]$) as the total sum. The belief bias parameter shows, based on each learner's sense of values, what is the value of each viewpoint and what is his/her tendency regarding the access to information, as well as how much importance each learner attaches to the viewpoint. When the parameter value of the viewpoint is large, the viewpoint is emphasized. The vector **b** is regularized, the sum total is assumed to be one, and the ratio of the emphasized viewpoints is shown as a parameter (Parameter of Belief Bias A). The vector **b** is calculated as follows:

$$b_j^{'} = w_j \sum_{i=1}^{m} d_{ij}, b = (1 / \sum_{i=1}^{n} b_i^{'})[b_1^{'}, b_2^{'}, ..., b_n^{'}] = [b_1, b_2, ..., b_n].$$

Next, the model for cultivating the inquiry-mindset in two learners, A and B, is shown in Figure 2 as an example. The information space that the learners A and B are accessing is shown as S_A and S_B in Fig. 2. The difference in the size of information space means the difference in the quantity, the degree of importance and the viewpoints of information accessed.

2.2 Flow of the Method for Cultivating the Inquiry-Mindset

The method for cultivating an inquiry-mindset proposed in this study consists of the following steps:

(a) Calculation of each learner's belief bias parameter,
(b) Group discussion referring to the presentation based on the calculation results of the belief bias parameter of each learner.

The following work is needed so that the method for cultivating the acquiring of an inquiry-mindset might progress appropriately:

- Calculations of the setting of each learner's tendency regarding access to information (information, degree of importance of information, and selection and evaluation of viewpoint of information),
- calculation of the setting of each learner's sense of values, and
- presentation of the belief bias parameter during group discussion.

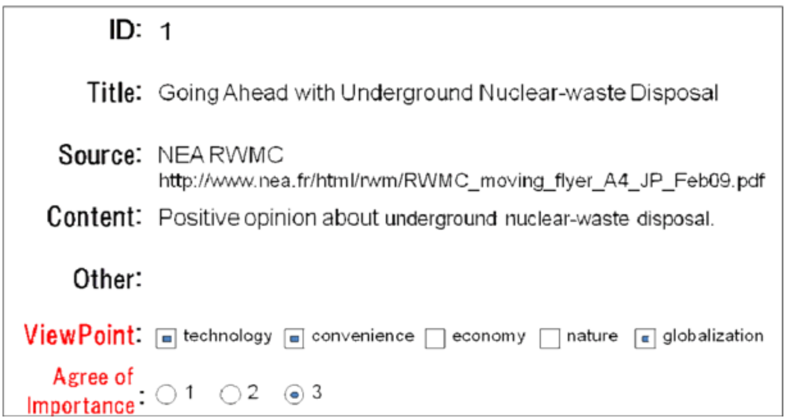


Fig. 3. Design for Information Evaluation Screen

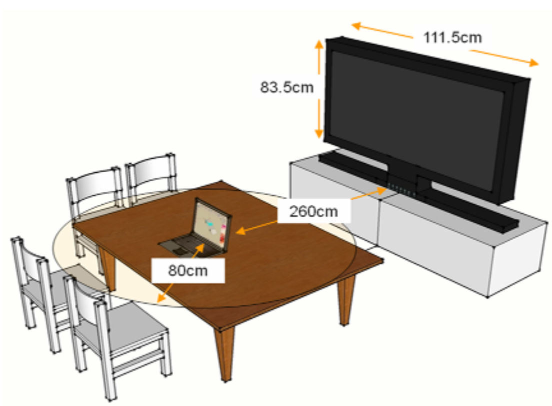


Fig. 4. Experimental Setting for Group Discussion

A supporting tool for the method of cultivating an inquiry-mindset is developed, because it is expected that the above-mentioned calculations and the presentation can be conducted more smoothly using a computer. The screen design for information input and belief bias parameter output during group discussion are shown in Figure 3 and Figure 4, respectively.

3 Evaluation Experiment

3.1 Purpose

The authors focus on information access bias based on the contents of information, importance degree of information and viewpoints of information, in order to quantify the belief bias and provide the learner with awareness about it. Therefore, the influence of these three elements on cultivating an inquiry-mindset is analyzed in the evaluation experiment. Because the place where the proposed method was applied and the measurement of the changes in the inquiry-mindset and the awareness to the belief bias are necessary for the analysis, the following conditions are required for the evaluation experiment:

- A place where evaluation of information and group discussion can be conducted,
- measurement of the inquiry-mindset before and after group discussion, and
- measurement of the awareness to belief bias before and after group discussion.

3.2 Method

Setting

- Participants: 38 people (undergraduate and graduate students, divided into ten groups of 3 or 4 persons).

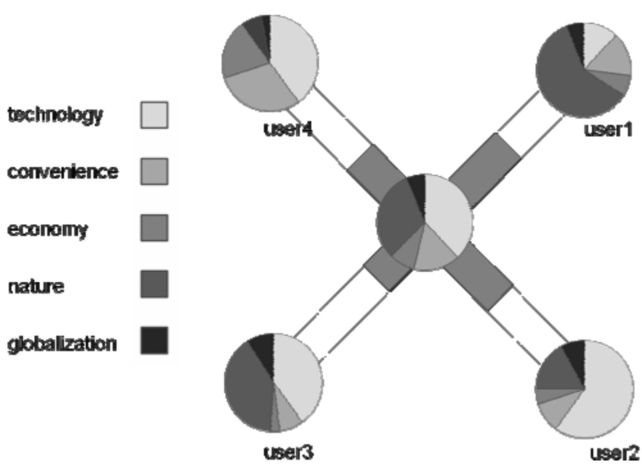


Fig. 5. Design for Belief Bias Parameters Output Screen

- Condition: "Approval or disapproval on underground disposal of high-level radioactive wastes" as a theme of controversial discussion.
- Five viewpoints (technology, convenience, economy, nature, and globalization) [11] are used for the evaluation of the viewpoint of information, known as a necessary framework for considering the environmental problems.

Procedures

- The flow of the task is shown below as the procedure of the experiment:

1. Setting one's sense of values (each participant)
2. Collection of information (each participant)
3. Sharing the collected information with the group and evaluating the degree of importance and viewpoints (each participant)
4. Group discussion

"1. Setting of one's sense of values", "2. Collection of information" and "3. Sharing and evaluation" take 65 minutes in total, and "4. Group discussion" takes 60 minutes.

Figure 5 shows the experimental setting for group discussion. Group members can use one personal computer, and each member can operate the screen displaying belief bias parameters of all group members (as seen in Fig.4).

Measurements

As for the measurements of the inquiry-mindset and the awareness to belief bias, the authors focused on changes in the situation before and after the experiments; a total of three measurements are conducted in the beginning of the experiment (pre-test), immediately before group discussion (middle-test), and immediately after group discussion (post-test). Questionnaires are used in pre-, middle- and post-tests.

Ten items (A1-A10) from the critical thinking attitude test (consisting of a total of 33 all items) [8] are used for the measurements of the inquiry-mindset, and evaluated on the Likert scale (seven point scale of +3: true to -3: not true). The test is executed three times (pre-, middle- and post-) and the changes in the inquiry-mindset are measured. The measurements of awareness to belief bias are divided into two parts: "Questions concerning information received in daily life (B)" and "Questions concerning the task of the experiment (C)." (B) includes three items (B1-B3) evaluated on the 7 point Likert scale. The tests are applied three times (pre-, post- and middle-) to evaluate the change to awareness of belief bias. (C) includes three items (C1-C3) evaluated on the 7 point Likert scale. The tests are applied once (middle) to evaluate the existence or non-existence of awareness to belief bias concerning the task of the experiment.

3.3 Results

The average points on the pre-, middle- and post-tests about the inquiry-mindset are shown in Figure 6. The average for the pre-test (1.5) and the post-test (1.7) has significantly improved ($t(37) = -3.2$, $p<0.01$). As for each item of A1-A10, the items A2, A3, A4, A6, A7, A8, A9 and A10 have higher average points in the post-test than in the pre-test. The contents of the items are related to the participants' motivation to study something new and the interest they expressed in different ideas. On the other hand, in the case of "A5: It is meaningful that we study how the foreigners think.", the average points in the post-test are lower than in the pre-test. The cause might be that the discussion theme is on domestic problems and the group members are all Japanese.

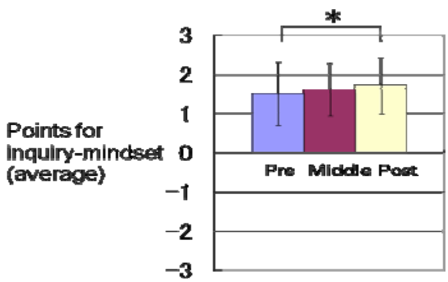


Fig. 6. Averaged Results of the Inquiry-mindset Evaluation (Evaluated on the 7 point Licker scale)

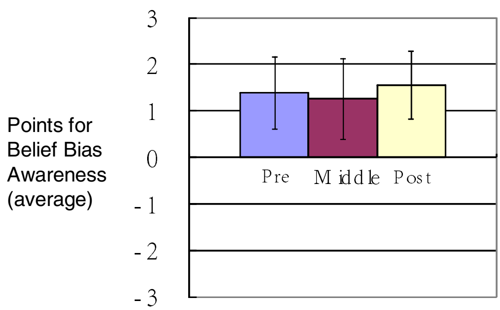


Fig. 7. Averaged Results of the Belief Bias Awareness Evaluation (B) (Evaluated on the 7 point Licker scale)

As for the correlation coefficients between the results about the inquiry-mindset and the awareness to belief bias, the latter are shown in Figure 7 and Figure 8. Fig. 7 shows the results pertaining to belief bias in the case of "Questions concerning information received in daily life (B)" (B1-B3) (seven point scale of +3: true to -3: not true) in the pre-, middle- and post-tests. Fig. 8 shows the results of belief bias on "Questions concerning the task of the experiment (C)" (C1-C3) (seven point scale of +3: true to -3: not true) in the middle-test.

The combinations with a positive correlation between inquiry-mindset and awareness to belief bias are A1 to A10 - B3, A4-B3, A10-B3, and A2-C1. On the other hand, the combination with a negative correlation is A9-B1.

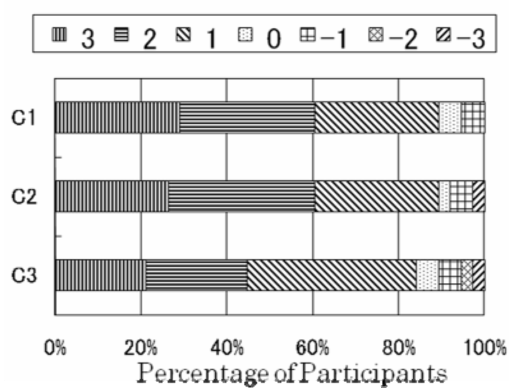


Fig. 8. Averaged Results of the Belief Bias Awareness Evaluation (C) (Evaluated on the 7 point Licker scale)

5 Conclusion

In this study, the authors aimed at proposing a new method for cultivating an inquiry-mindset focusing on the awareness to belief bias, in an attempt to solve problems that have not been tackled in previous research. The authors proposed a method for cultivating an inquiry-mindset using the belief bias parameter, the effectiveness of which was verified. The results suggested that it is possible for access to information and awareness to the bias of the viewpoint of information to influence the improvement of the inquiry-mindset.

Future challenges refer to developing the proposed method based on the results of the experiments. One of the practical challenges would be to use this method in universities and evaluate and analyze its efficiency in practice.

Acknowledgement. This work was supported by Grant-in-Aid for Scientific Research (C) 21500874.

References

1. Weinberg, A.M.: Science and Trans-scienceMinerva, Tokyo (1972)
2. Michita, T.: Critical Thinking of University Students in Reading Non-Academic Materials: Attitude and Ability Differences in Relation to Academic Level and Major. The Japanese Journal of Educational Psychology 49(1), 41–49 (2001) (in Japanese)
3. Hirayama, R.: Effect of Critical Thinking Disposition on Interpretation of Controversial Issues: Evaluating Evidences and Drawing Conclusions. The Japanese Journal of Educational Psychology 52(2), 186–198 (2004) (in Japanese)
4. Inaba, M.: Enhancing Critical Thinking Disposition among College Students through the Collaborative Construction of Learning Materials on the CSCL Environment. Policy Science 14(3), 13–24 (2007) (in Japanese)

5. Aoyagi, S., Itami, Y., Shimoda, H., Ishii, H., Tomie, H., Kitagawa, K.: Experimental evaluation of a Curriculum using Educational Debate System for fostering Critical Thinking Attitude. Human Interface Society 10(5), 45–50 (2008) (in Japanese)
6. Markovits, H., Nantel, G.: The belief-bias effect in the production and evaluation of logical conclusions. Memory and Cognition 17, 11–17 (1989)
7. Herman, E.S.: The Propaganda Model: a retrospective. Journalism Studies 1, 1–14 (2000)
8. Gentzkow, M., Shapiro, J.M.: What drives media slant?: Evidence from U.S. daily newspapers (No. Working Paper No. 12707). National Bureau of Economic Researcher, Cambridge (2006)
9. Kai, K., Yamamoto, H., Osato, M., Shiino, T., Kato, M., Sawayama, Y.: Understanding Student's Utterance in ICAI System for English Conversation. Information Processing Society of Japan 1, 22–23 (1989) (in Japanese)
10. Inami, I.: Smart Organization Decision Making. Chuokoron-shinsha, Inc., Tokyo (1989)
11. Sajima, T., Takayama, T., Yamashita, H.: Basic & fundamental learning of resources, energy and environment. Kokudosha, Co., Ltd., Tokyo (2000) (in Japanese)

Distance Education at the Graduate Level: A Viable Alternative?

Brian M. Jones[1], Andrea Everard[2], and Scott McCoy[3]

[1] Tennessee Technological University, College of Business
Cookeville, TN 38505, USA
bjones@tntech.edu
[2] University of Delaware, Lerner College of Business and Economics
Newark, DE 19716, USA
aeverard@udel.edu
[3] The College of William and Mary, Mason School of Business
Williamsburg, VA 23187, USA
scott.mccoy@mason.wm.edu

Abstract. A prior comparative study examining learning outcomes between traditional classroom and web-based education at the graduate level provided little evidence of differences in success measures between delivery methods [3]. This research explores underlying factors that may explain why little difference was found.

Keywords: distance education, on-line learning, computer based learning, education

1 Introduction

Distance education is an alternative educational model to the traditional classroom. Distance education is continually evolving. Because geographical and temporal separation is no longer considered an obstacle to course delivery thanks to the Internet and other information technologies, an increasing number of individuals are now able to partake in this new educational opportunity.

This study is a follow-up to previous research conducted on the differences in learning outcomes between traditional and web-based course delivery for a required MIS course at the graduate level [3]. The original longitudinal research concluded that there were no significant differences in learning outcomes as measured by exam grades and final course grades of graduate students exposed to traditional and web-based educational settings. This paper extends the previous work by identifying student characteristics that may explain why no statistically significant results were noted and hence make web-based education a viable and useful alternative to traditional course delivery for certain student populations.

2 Prior Research

There has been an increase in online education programs with the introduction of the Internet and the abundance of educational applications [5]. Online learning

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 58–60, 2011.

management systems enable students to control when and where they learn [4]. Distance education is often considered a more convenient alternative especially for non-traditional students given its ability to provide 'anytime', 'anywhere' education. Although not all types of students find online education attractive or suitable, often noted benefits include time and location flexibility [1] [2] [7], and having a sense of control over the learning environment [6].

3 Proposed Research

The results of the original research study [3] which compared and contrasted traditional on-campus learning and web-based education were different than anticipated. Because of this, the current study attempts to determine the underlying factors explaining the unexpected results of the initial study.

In the original research study, learning outcomes (final exam grade and final course grade) were measured for an introductory MIS graduate course that was delivered through a traditional method of on-campus classroom lecture and discussion and a web-based learning environment using iLearn and WebCt technologies. Except for the delivery mode, the courses were identical: the same faculty member taught, administered and oversaw all classes whether on-campus or provided through the web-based program, using the same lecture notes, tools, assignments, and tests. In the web-based course, lectures were recorded using voice with MS Powerpoint presentations for some modules and actual video-taped segments for others. Both classes included a participation requirement and both were encouraged to develop discussion topics and to foster them using the tools available through iLearn and WebCT. No significant differences in learning outcomes were reported.

At the time of the data collection for the initial study, data was also collected on other variables such as employment status, number of children that reside in the house, family commitments, age and gender and computer efficacy in a naturally-occurring educational environment. This additional data will provide the needed insight into how these student characteristics affect the viability of web-based education.

4 Summary and Conclusion

These results will be beneficial to both instructors and students. Students will be able to make better informed decisions as to whether they are suitable candidates for web-based education. Instructors will be able to evaluate whether web-based education can deliver a similar learning experience as a traditional on-campus environment.

References

[1] Glenn, A.: A comparison of distance learning and traditional learning environments. ERIC Document Reproduction Service, ED457778 (2001)

[2] Johnson, S.D., Aragon, S.R., Shaik, N., Palma-Rivas, N.: Comparative Analysis of Learner Satisfaction and Learning Outcomes in Online and Face-to-Face Learning Environments. Journal of Interactive Learning Research 11(1), 29–49 (2000)

[3] Jones, B., Everard, A.: Web-Based vs. Traditional Education: Does Distance Education at the Graduate Level Deserve a Failing Grade? In: Proceedings of AMCIS 2008, Toronto, Canada (2008)

[4] Lockhart, J.: Learning to fit in with real world – online and distance education. The Australian (April 26, 2006)

[5] Natriello, G.: Modest changes, revolutionary possibilities: Distance learning and the future of education. Teachers College Record 107(8), 1885–1904 (2005)

[6] Petracchi, H.: Distance education: What do our students tell us? Research on Social Work Practice 10(3), 362 (2000)

[7] Rosenbaum, D.B.: E-learning beckons busy professionals: Electronic education offers anywhere, anytime flexibility...but not without problems. ENR 246(21), 38–42 (2001)

Creating a New Context for Activity in Blended Learning Environments: Engaging the Twitchy Fingers

Jayne Klenner-Moore

King's College, Department of Computers & Information Systems, USA
JayneMoore@kings.edu

Abstract. Millennial students are used to environments that bring the information to them. iPads, iPods and Web 2.0 technologies that give today's professors the tools to interact at many levels with students. Students want to be able to do things after class time and between *World of Warcraft* games. This paper provides a case study of melding technology and pedagogy with learning theory to get students engaged in the learning in computer classrooms. Students way of interacting with information has changed, the way we deliver and support this exchange needs to as well. This paper will include a look at how students interact with technology and defines the types of learning objects and opportunities provided by the technology. Two levels of students were examined, middle school Hispanic computer skills learners and undergraduate students in a project management class. This paper reports on observations made during these classes.

1 Introduction

In the summer of 2011 DyKnow™ was used in a computer class for middle-school level students of the Hispanic Outreach Program. The software was part of a grant that was given to the program along with tablet notebooks. The goal was to increase student interaction with learning. This population traditionally does not have access to computers and education past high school is not encouraged in the home. The Outreach program states, "The King's College Hispanic Outreach Program's focus is to create an organized framework for the purpose of building strong relationships between the Latino Community and King's College, as well as assist in addressing the needs of the Latino community and their families." (1) Engaging students in the learning environment was crucial.

The interest and use of this technology by the middle-school students encouraged the introduction of the software to a college classroom. Goals of these two "pilot studies" were; to observe student interaction with the software, observe access to the notebooks outside of class time, and to see if students were more involved with the learning during class time.

Middle-school level students used the software during class time only while the college level students were able to hold class both in the class and remotely with the software. All students were able to access the "notebooks" and recorded lessons from the DyKnow server during the term of the instruction. The terms were three weeks for the summer middle-school students and one semester for the college students.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 61–67, 2011.

2 Learning Contexts

The well designed instructional environment employs technologies that enable students to interact with the learning experience in ways that are relevant and enhance learning. Supporting software programs provide for a storage and retrieval area of pertinent course information. This allows the student to review and interact with the materials again over time.

This also changes the context of the learning experience. In this case, the context moves from human-to-human interaction to one of computer-mediated action that includes not only the computer as tool and human as actor but includes the larger “world” of the class. The students are no longer interacting only with the computer, but with each other in a goal-directed activity of learning. Kaptelinin states, “According to activity theory, the hierarchical organization of human computer interaction is determined by its emeddedness into the hierarchical structure of human activity that mediates the user’s interaction with reality” (2). Further he states, “Meaningful, goal-directed activities constitute the context for both mental processes and external actions.”

In an article from Edutopia, Mac Prensky quotes students from one of his panels on what education should be, “There is so much difference between how students think and how teachers think," offered a female student in Florida. A young man commented, "You think of technology as a tool. We think of it as a foundation -- it's at the basis of everything we do” (3). This feeling calls for a change in pedagogy and a call for using technology in the context of the classroom.

Blended learning environments use traditional classroom interaction along with web-based interactions. This allows the classroom to extend beyond the time and place of a traditional classroom. This paper looks at once such technology that can be used to design participatory learning experiences and to extend the learning interaction time for students. The technology is DyKnow. DyKnow is a classroom management software technology that enables collaboration and active participation in the learning. This technology allows students to review the actual class experience through the web anytime, anywhere.

DyKnow provides web storage of class notes and “notebooks” created by the students and the teacher in a fashion similar to old fashioned note taking. During class time students each have a view on their individual computer screen that is guided by the instructor. Outside of class students can access their saved “notebooks” to review their own notes and those made by the instructor. The instructor may also choose to record the session so students can hear the discussion when doing a review.

Many of today’s classrooms come equipped with either wireless connection for students to access the net or computers for students to use during class time. Unless the context in the classroom is designed for learning students are often found twitching their fingers and going to places like Facebook and email during class time. The irresistible tug of technology leaves them disinterested in the lecture/note taking portion of class. However, if there is a purpose for their fingers to be active in the class the instructor can now work more efficiently at engaging the learner.

3 What Were We Looking For?

Both classes used to examine the DyKnow technology were taught in classrooms with computers at each student seat. One mission of using this software was to engage the students in the actual activity and content of the courses. There is a lot of research being conducted on the millennial student and how she learns due to the influence and omnipresence of technology in her life. Having the DyKnow software available for student use in the classroom it was hoped that students would engage more directly and in collaboration with the learning.

Many students do not take notes in class anymore; they come in as very passive learners expecting to be entertained. By providing a dynamic workspace on the classroom computers with the DyKnow technology students were able to take notes in a manner consistent with their daily practice of texting or utilizing a computer. This afforded a working space that used computer technology along with active participation in things like note taking and thinking-sharing.

The following sections discuss the instructor observations and student feedback of these studies.

3.1 Limitations to the Studies

While the observations and student feedback regarding this method of interaction were promising the small sample size was not sufficient to make generalizations to a larger population. It was encouraging enough however to proceed with introduction of the pedagogy to colleagues and develop future research methods to better analyze learning outcomes. Larger groups and the inclusion of a control group for further study would be optimal. Additionally subsequent semesters afforded a look at using this technology with visually impaired students.

4 Active Learning with Technology as Part of Blended Learning

4.1 Active Learning

Berque, Byers and Myers talk about a “pedagogical technique that is sometimes referred to as the upside down classroom” (4) in this classroom students come prepared to be active and learn outside the classroom as well as during class.

“Learning is a willful, intentional, active, conscious, constructive practice that included reciprocal intention-action-reflection activities” (4).

4.2 Blended Learning

Blended learning is described as “a hybrid of classroom and online learning that includes some of the conveniences of online courses without the complete loss of face-to-face contact” (5). Blended learning environments can be designed along a continuum of pedagogy and technology inclusion.

Today’s technologies allow more opportunities for students to access live classes and retrieve information on course concepts at their own pace and on demand. Models for the development of open learning environments (7) provide a structure for the practical application of technology in creating accessible learning environments that are learner-centered in nature.

Moller, Prestra, et al (8), talk about organic knowledge building, which focuses on the individual learner within a learning environment in relation to asynchronous learning environments.

5 Key Differences between Studies

5.1 Reason for Learning Is Different

"Learning is a willful, intentional, active, conscious, constructive practice that included reciprocal intention-action-reflection activities" (9). The introduction of DyKnow software to the classroom gave students a focused action plane in which to engage with course content and develop constructive practice with other students engaged in the same activities.

The ability to review course content at will from anywhere permitted students to interact longer and more frequently with the material. This ability reinforced the learning. Students remembered more and in the case of the middle school children improved English language skills such as spelling and writing.

5.2 Teacher Satisfaction Using Active Learning

The level of excitement due to the interaction made lesson planning and class interaction more satisfying. One day former students of the program, now in college, came in to work with the middle school students. These students were able to work with DyKnow on the layout activity for the newsletter that was created.

It is imperative, as with any change, that there is buy-in by the stakeholders. If using this active learning approach in traditional and blended learning environments increases teacher satisfaction it will be adapted more readily.

(Dyknow white paper states)By breaking the teacher-centered discussion into small chunks, active learning strategies rely on fostering student engagement (10). This paper which discusses best practices for active learning also points out, "an additional benefit of many active learning is the immediate feedback they provide to teachers". By using the polling feature of DyKnow the teacher can ask immediately for feedback on any item in the lesson. This allows for teaching moments or areas of further research.

6 Main Advantages Learned from These Studies

6.1 Engaging the Twitchy Fingers – Computer Mediated Activity

Students were able to take notes along with the lecture right on the screen where the activity or learning was taking place. They could highlight, tap notes, chat and share screens with the teacher and the class at large. This activity enhanced their focus on the lesson material while engaging their action toward the goal of learning. Their energy was harnessed in learning activities and not just computer browsing.

6.2 Peer Sharing of Questions, Comments, Notes, etc

One of the features of DyKnow™ is that it allows for peer sharing of questions, comments and notes via the interactive interface. Additionally, when an interactive white board is used with the class the teacher's annotations are visible to the students and their individual panels can be viewed on screen and on the whiteboard.

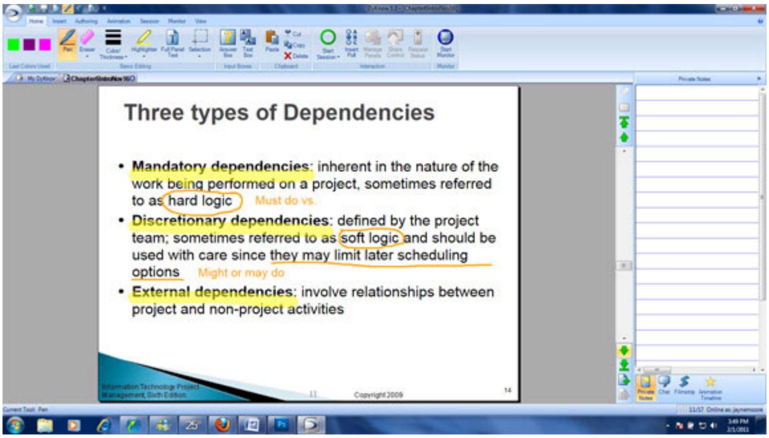


Fig. 1. Screen Shot of Teacher Presentation Notes Using DyKnow

The example shown here is the screen from the teacher's station. The markings are from the teacher and can be viewed in each student's notebook on their screens. This is then saved to the student notebooks which can be reviewed later.

Each student panel can be viewed and gathered by the instructor and then shared with the class or marked with feedback and returned to the student.

7 Twitchy Fingers and the Millennial Learner

Ipods, iPads and computer games have given rise to students who engage the world through their fingertips. These students text each other during class and are used to accessing information at their fingers. Touch screens and keypads have put fingers in constant motion all day long this leads to the twitchy fingers syndrome. This affects student behavior while sitting in the classroom, particularly a computer classroom.

In order to engage the learner it is necessary to engage their twitchy fingers in some goal-directed action or learning activity.

Millennial students are used to environments that bring the information to them. iPads, iPods and Web 2.0 technologies give today's professors the tools to interact at many levels with students during the semester and perhaps even longer. Learning does not have to stop when the proverbial class bell rings. Students want to be able to do things after class time and between *World of Warcraft* games. Creating learning environments that allow the student to review the class, access personal notes and review relevant materials can actually elicit better learning habits among this learning group.

8 Scaffolding Learning with DyKnow Interface

Land and Hannafin make the point that, "…a lack of external support is mistaken for student-centered learning despite the absence of needed scaffolding" (11). DyKnow and active learning contexts indeed scaffold the learning as they provide teacher supported, material supported and peer-supported tools to enable the learner to approach the learning with the appropriate scaffolds via the materials and peer interaction that is guided by the instructor.

Each of these experiences and tools provides multiple interactions with the learning at several levels over time.

9 Future Research

Due to the successful feedback from students and the excitement generated in the classroom for the instructor the licensing grant for the use of DyKnow™ at the college was extended. A closer look will be taken with regard to learning outcomes. Additionally, we will be able to use this again with middle and high school students over the summer. The interest shown this method by students allows us to now focus on learning outcomes and motivation to learn with computer mediated tools.

References

1. Shively, B.: home. Hispanic Outreach, http://departments.kings.edu/hispanicoutreachprogram/home.html (cited: November 2, 2011)
2. Kaptelinin, V.: Computer-Mediated Activity: Functional Organs in Social and Developmental Contexts. In: Nardi, B.A. (ed.), pp. 45–68. MIT Press, Cambridge (1996)
3. Prensky, M.: The 21st-Century Digital Learner. Edutopia (May 22, 2008), http://www.edutopia.org/ikids (accessed on February 2, 2011)
4. Burque, D., Byers, C., Myers, A.: Turning The Classroom Upside Down Using Tablet PCs And DyKnow Ink And Audio Tools. In: Workshop on the Impact of Pen-Based Technology on Education (2008)
5. Jonassen, D.H., Land, S.M.: Preface. Theoretical Foundations of Learning Environments. Lawrence Erlbaum Associates, Mahwah, pp. iii – ix (2000)
6. Jordan, M., Rovai, A.P., Hope.: Blended Learning and Sense of Community: A comparative analysis with traditional and fully online graduate courses. The International Review of Research in Open and Distance Learning (August 2004), http://www.irrodl.org/index.php/irrodl/article/viewArticle/192/274.7 (cited February 2, 2010)

7. Hannafin, M., Land, S., Oliver, K.: Open learning environments: Foundations, methods and models. In: Reigeluth, C. (ed.) Instructional Design Theories and Models, pp. 115–140. Lawrence Erlbaum Associates, Mahwah (1999)
8. Moller, L., Prestera, G.E., Harvey, D., Downs-Keller, M., McCausland, J.-A.: Creating an Organic Knowledge-Building Environment within an Asynchronous Distributed Learning Context. Quarterly Review of Distance Education 3(1), 47–58 (2002)
9. Jonassen, D.H., Land, S.M.: Preface.Theoretical Foundations of Learning Environments, p. 11. Lawrence Erlbaum Associates, Mahwah (2000)
10. Emerging Active Learning: Best Practices and Supporting Technologies. White paper
11. Hannafin, M., Land, S., Oliver, K.: Open learning environments: Foundations, methods and models. In: Reigeluth, C. (ed.) Instructional Design Theories and Models, p. 11. Lawrence Erlbaum Associates, Mahwah (1999)

Haptically Enhanced User Interface to Support Science Learning of Visually Impaired

Yueqing Li, Steve Johnson, and Chang Nam

University of Arkansas, Fayetteville, AR, United States
{yxl002,sajohns,cnam}@uark.edu

Abstract. The primary purpose of this study is to evaluate the overall quality of haptic user interfaces designed to support various science learning activities in order to obtain usability and performance data. The result showed that haptic interface could significantly affect impaired students' user performance. Audio & Tactile interface can produce significantly better user performance than the Tactile interface. Meanwhile, the learning effect could be greatly enhanced and students enjoyed the hands-on experience very much. This study should provide invaluable empirical data and some insight for the future research.

Keywords: haptic, tactile, visual impairment, interface.

1 Introduction

According to National Health Interview Survey in 2008, more than 25.2 million Americans reported experiencing vision loss [1]. Statistics from Braille Institute showed that vision problems affect 5% (about 5 million) of school-age children, ages 3-5 and 25% (12.1 million) of school-age children, ages 6-17 [2]. Although visually impaired, they have the right to receive the same science education as other people with normal sight. However, one big challenge for visually impaired students is that: most of the materials (e.g., textbooks, graphics, etc) for science learning are still visually based. According to Patton & Braithwaite [3], almost 90% of science teachers who teach the visual impaired students teach science class mainly based on textbook. White [4] argued that science textbooks generally present science concepts in the most abstract formats and mathematical models. But since visually impaired students' main sensory channels are tactile and auditory, textbooks are, needless to say, insufficient in meeting such needs. Meanwhile, visually impaired students' preconceptions about natural phenomena may differ from the accepted scientific concepts [5]. Considering that most science concepts are intrinsically abstract, visually impaired learners often have difficulty connecting abstract science concepts to sensory experience-based knowledge [6, 7].

To help visually impaired students learn more efficiently and directly, haptic technology has been increasingly applied across multiple domains [8-10]. Haptics utilizes tactile feedback to manipulate a variety of touch-based sensorial experiences. Different methods can be employed to realize haptics, such as forces, vibrations, and motions [11]. Some researchers propose that haptic perception (e.g., force,

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 68–76, 2011.

vibrotactile and thermal), combined with audio information, can improve visually impaired students' ability to understand scientific concepts [12, 13]. Moreover, visually impaired students can better acquire information through haptic sensations [14]. With sensorial feedback, visually impaired students can conceptualize and retain scientific mental models more easily [15]. Even more importantly, hands-on haptic-based science learning experiences instill greater confidence and increase critical problem solving [9]. Despite recent advances in haptic research, research opportunity still exists, especially in the usability analysis of haptic applications.

The primary objective of this study is to evaluate the overall quality of haptic interfaces designed to support various science learning activities (e.g., menu selection, navigation and recognition of molecular structure and force, etc.) in order to obtain usability and performance data, as well as to refine the design guidelines for haptically enhanced science learning systems.

2 Haptic Interfaces

The Molecular Properties Module (MPM) is a haptically enhanced science learning system that provides visual, haptic, and auditory feedback for students with visual impairments to learn molecular concepts such as molecular structure and intramolecular force. MPM was developed to facilitate haptic science learning for students with visual impairments and supports two key tasks: molecular structure recognition and intramolecular force recognition.

2.1 Molecular Structure Recognition

Molecular Structure Recognition supports Tactile, Audio, and Audio & Tactile interfaces. Each haptic interface supports the display of three molecules (CO_2, H_2O, and BF_3) as Two-Dimensional (2D) Ball & Stick molecular models (Fig. 1). Each molecule supports haptic force feedback to enable users to trace around the edges of the model using a haptic device.

The tactile interface adds tactile feedback to a group of atoms within the model. This tactile feedback allows a user to further distinguish one atom from another within a molecular model. For example, H_2O contains three atoms – one Oxygen (O) and two Hydrogen (H) atoms – and the connecting sticks. When the user touches the Oxygen atom, a smooth tactile effect is rendered; when the user touches either

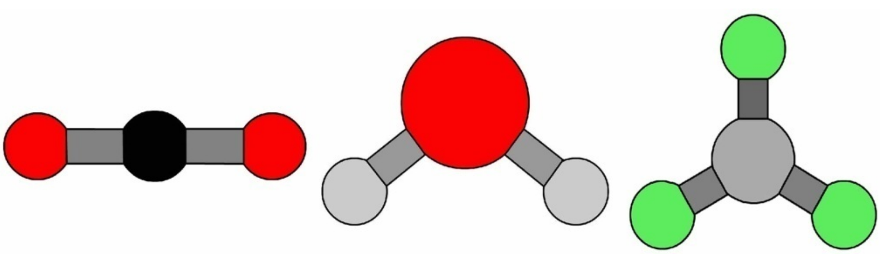


Fig. 1. Molecular structure recognition: *left (CO_2)*, *middle (H2O)*, *right (BF3)*

Hydrogen atom, a bumpy tactile effect is rendered; when the user touches any connecting sticks, a sandpaper tactile effect is rendered.

The Audio interface adds audio feedback to each atom within the haptic molecular model. A pitch is played every time the user's cursor comes into contact with an atom. As with the previous H_2O example, when the user touches the Oxygen atom, a medium-pitched audio effect is played; when the user touches either Hydrogen atom, a high-pitched audio effect is played; when the user touches any connecting sticks a low-pitched audio effect is played.

The Audio & Tactile interface combines both audio and tactile modalities described above into one interface type.

2.2 Intramolecular Force Recognition

Intramolecular Force Recognition supports two molecules (CO_2 and CS_2) as molecular spring models. Each model allows the user to manipulate (i.e. grab) the atoms surrounding the central atom. For example, the user could manipulate the Oxygen (O) atoms in CO_2 and Sulfur (S) atoms in CS_2. Gravity Wells provide a haptic sensation of the delicate balance of attractive and repulsive forces between atoms in the molecule. A gravity well is a haptic tool that automatically snaps the user into the center of an area when the user's cursor is within a predefined pull radius.

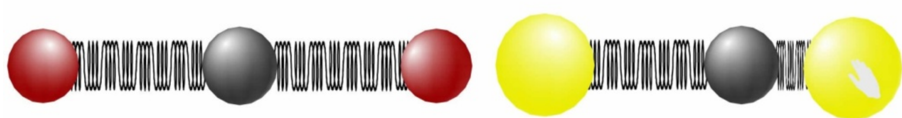


Fig. 2. Spring model for intramolecular force recognition: *left (CO_2), right (CS_2)*

3 Methods

3.1 Participants

Twelve participants were recruited from Arkansas School for the Blind in Little Rock, AR. Participants were recruited by instructors at the school and were monetarily compensated for their voluntary participation. There were 5 female and 7 male participants whose mean (M) age was 14.2 years (Standard Deviation, SD = 2.0). All participants had little or no experience with haptic interfaces and haptic devices.

3.2 Apparatus

Hardware includes a Dell PC with a 3.4 GHz Pentium R and 1.0GB of RAM, up to 2 Novint Falcon haptic devices. Software includes Adobe Flash CS3 for visual rendering and a C/C++ Novint SDK for haptic rendering to develop the software. Haptic features were multithreaded, graphics were rendered at 60 Hz, and haptic control was updated at a rate of 1 kHz.

To evaluate the effect of haptic interface, multiple interface types, and haptic interaction methods were developed for students with visual impairments. This study supported 2 key tasks and 4 interface types, as shown in Table 1.

Haptic Interface Evaluation Surveys were designed. They consisted of two questionnaires, one for each key task. Each questionnaire evaluated user preference in regards to the interface types supported by a particular task as well as user comments for each individual interface type supported by each task.

NASA Task Load Index (TLX) was used to measure the workload that participants experienced. It contains six subscales measuring mental demands, physical demands, temporal demands, performance, effort, and frustration. All items were rated on a 10-point scale.

3.3 Independent Variables

The Molecular Structure Recognition task contained only one independent variable: haptic interface. Three haptic interfaces were provided: Tactile interface, Audio interface, and Tactile & Audio interface. In the Intramolecular Force Recognition task, no independent variables were manipulated.

3.4 Dependent Variables

Dependent variables include user performance measurement and user behavior measurement. Key task questionnaires were also utilized to determine interface preference as well as suggestions, improvements to each interface type for the 2 key tasks. Performance Measurements includes workload, task completion time, and Pass/Fail marks (success rate).

To measure user behavior, a Cursor Trajectory Management System (CTMS) was developed in order to collect and store all user cursor behavior throughout each scenario. The CTMS measures data at an interval of approximately 20 milliseconds. A Position Data Analysis (PDA) was conducted on the collected user behavior data. The user's cursor positional data was subdivided into zones to quantify proportions of the user's total cursor activity/behavior.

- **Error Region (ER):** Region(s) surrounding *Stability Region*(*s*) or inside a 2D haptic object. Signifies that the user is not near to any haptic object and is considered poor user behavior.
- **Stability Region (SR):** Region(s) surrounding Trace Region(s). Signifies that the user is very near to a haptic object, within about 10-20 pixels, and is considered good user behavior.
- **Trace Region (TR):** Region(s) closest to haptic object(s). Signifies user interaction with a haptic object and is considered perfect user behavior.

Positional Data were aggregated and proportionally divided to describe the user's total actions for a particular scenario:

- **Error Region Proportion (ERP):** Proportion of total user activity within the ER region(s) of a particular interface.
- **Stability Region Proportion (SRP):** Proportion of total user activity within the SR region(s) of a particular interface.

- **Trace Region Proportion (TRP):** Proportion of total user activity within the TR region(s) of a particular interface.

3.5 Procedure

Prior to the experiment, each participant was required to complete a Consent Form as well as a Demographics Form. Next, each participant was required to complete a Haptic Training Session, which could provide a foundational understanding and sensibility of the Novint Falcon haptic device as well as haptic virtual environments. After the training session, the participant then participated in two uniquely designed key tasks. The sequence of key tasks, haptic interfaces, and scenarios were counterbalanced to remove the learning effect as much as possible. Each scenario was conducted as follows: A participant read a scenario description outlining what haptic objects would be present as well as the scenario goal. During this time, a participant could ask any relevant questions – as long as it did not reveal sensitive information regarding how to go about completing the scenario goal. Participants were told that each scenario had no time limit, though a 10 minute cut off point was enacted if necessary. Finally, participants were made aware that if they felt that they had satisfied a particular scenario's ending condition prior to reaching the time limit, they could say "Done." Upon each scenario's conclusion, participants were asked to complete a NASA TLX questionnaire regarding the completed scenario's cognitive workload. Additionally, upon each key task's conclusion, participants were asked to complete a User Preference Questionnaire to obtain user preference and comments in regards to each interface type within a key task. Table 1 provides a detailed description of each Key Task.

Table 1. Key task description

Task	Description	Pass Condition
Molecular Structure Recognition	To feel a Ball and Stick Model of a molecule. The model has 3 interfaces. When finished, please say "I'm Done" and draw the molecule you just felt.	The user must correctly draw the molecular model (# of atoms, and geometry).
Intramolecular Force Recognition	To feel the intramolecular force between atoms within a molecule for 2 molecules. Locate an atom at first, then use the ACTION button to grab and move the atom to feel the force. When finished, please say "I'm Done" and tell the result of force comparison.	The user must answer that CS2 has a greater intramolecular force

4 Result

4.1 User Performance

In the Molecular Structure Recognition task, the average success rate was 72.2%. ANOVAs were conducted to determine the effect of haptic interface on success rate, workload and task completion time. For Success Rate, significant effect of haptic interface was found ($F_{2, 22} = 4.53$, $p < 0.05$). Further analyses showed that success rate in the Tactile & Audio interface (M = 92%, SD = 29%) was significantly higher than that in Tactile interface (M = 42%, SD = 51%). However, there was no significant difference in success rate between Tactile interface and Audio interface (M = 75%, SD = 45%), or Audio interface and Tactile & Audio interface.

No significant effect of haptic interface was found for workload and task completion time in the Molecular Structure Recognition task. However, the mean of workload in Tactile & Audio interface (M = 3.72, SD = 1.54) was higher than that in Audio interface (M = 3.42, SD = 1.36), which was higher than that in Tactile interface (M = 3.26, SD = 1.47). The mean of Task completion time in Tactile & Audio interface (M = 72.89, SD = 40.02) was shorter than that in Audio interface (M = 82.19, SD = 42.22), which was shorter than that in Tactile interface (M = 84.04, SD = 47.28). In the Intramolecular Force Recognition task, participants reached an average success rate of 66.7%.

4.2 Behavior Performance

ANOVAs were conducted to determine the effect of haptic interface on Error Region Proportion (ERP), Stability Region Proportion (SRP) and Trace Region Proportion (TRP) in the Molecular Structure Recognition task. No significant effect of haptic interface was found for behavior performance. However, ERP in Tactile & Audio interface (M = 36.79%, SD = 17.86%) was smaller than that in Tactile interface (M = 41.29%, SD = 24.19%) and Audio interface (M = 41.16%, SD = 21.54%) on average. TRP in Tactile & Audio interface (M = 34.37%, SD = 12.83%) was bigger than that in Tactile interface (M = 30.49%, SD = 16.92%) and Audio interface (M = 30.64%, SD = 15.04%) on average. But SRP were almost the same in the three interfaces: Tactile interface (M = 28.22%, SD = 8.84%), Audio interface (M = 28.2%, SD = 8.74%), and Tactile & Audio interface (M = 28.84%, SD = 8.28%). In the Intramolecular Force Recognition task, participants finished the task with an average 22.54% ERP, 62.47% SRP, and 14.99% TRP.

5 Discussion

Since there were three haptic interfaces in the Molecular Structure Recognition task and only one in the Intramolecular Force Recognition task, the two tasks would be discussed separately. For the Molecular Structure Recognition task, the focus is the effect of the haptic interface. For the Intramolecular Force Recognition task, the focus is the user performance and the applicability of the haptic application.

5.1 Molecular Structure Recognition

Effect of Haptic Interface on User Performance. Result showed that haptic interface had a significant effect on success rate, which was significantly higher in Tactile & Audio interface (M = 92%) than that in the Tactile interface (M = 42%). Success rate in Tactile & Audio interface also had a larger average value than that in Audio interface (M = 75%), but it is not significant.

This result indicates that the Tactile & Audio interface provides more information for user with visual impairments. The Tactile & Audio interface, in addition to haptic force feedback, provides tactile and audio feedback. The combination of force, tactile, and auditory feedback must enhance the user's conception of the molecular structure, leading to better performance. The relevant shorter task completion time in Tactile & Audio interface (M = 72.89) compared to the Tactile (M = 84.04) and Audio interfaces (M = 82.19) further supports that the Tactile & Audio interface produces better user performance.

User Preference Questionnaires indicated that 12 out of 14 participants preferred the Tactile & Audio interface. User comments indicated that the combination of tactile and auditory feedback facilitate ease of use when tracing molecules, feeling and recognizing individual atoms, and visualizing an entire molecular shape.

However, more information also means more cognitive workload. The result showed that participants had relevantly higher workload in the Tactile & Audio interface (M = 3.72) than the Tactile (M = 3.26) and Audio interfaces (M = 3.42). Two participants that did not prefer the Tactile & Audio interface commented that the combination of force, tactile, and audio feedback caused confusion and disorientation during the task. Such a level of cognitive workload can affect the user's concentration and can decrease usability and potentially limit haptic applications. As such, if users have difficulty processing multiple sensorial feedbacks, resulting confusion or disorientation can produce negative effects. For example, higher levels of cognitive workload can cause users to easily tire and lose concentration, limiting the applications effectiveness. Therefore, to create a user-friendly and accessible haptic application, more research is needed to establish the relationship between cognitive workload and acceptable sensorial modalities within a haptic application.

Effect of Haptic interface on User Behavior. Haptic interface showed no effect on user behavior. Although interface type could significantly affect user performance, the difference was not large enough to affect user behavior. However, on average, user behavior showed the smallest ERP and largest TRP in the Tactile & Audio interface among the three types of haptic interfaces. Because TRP and SRP represent good and perfect user behavior, relevantly larger TRP and smaller ERP signifies that the Tactile & Audio interface conditions produce optimal user behavior per task instruction, supporting the hypothesis that multiple modalities can lead to greater interface usability. However, since the effect is not significant, more research is needed.

Haptic Interface Usability Limitations. Although most participants preferred the Tactile & Audio interface, some participants commented that the audio was not systematically designed and that auditory pitch information was, at times, confusing. Considering that the visually impaired may have greater sensitivity and sensibility in receiving and processing auditory information, more research is needed to define user-friendly audio design for those with visual impairments.

Some participants also commented on the difficulty of navigating and locating haptic objects within some interfaces. More specifically, it seems as though these participants had difficulty in developing accurate mental models of the two dimensional environments in order to correctly navigate and locate haptic objects, as well as determine their haptic cursor location relative to the haptic objects within the environment. To better meet the requirements of the visually impaired, a more intuitive navigation should be designed to ensure that users can build accurate mental models of the 2D environment. As a possible solution, perhaps one button on the haptic device could be utilized to automatically return users to a specified area within the haptic environment in the event that the user is lost.

5.2 Intramolecular Force Recognition

In the Intramolecular Force Recognition task, participants finished the task with an average success rate of 66.7% (8 out of 12). The average ERP, SRP and TRP were 22.54%, 62.47% and 14.99%, respectively. Result showed that most participants could perform the task with satisfactory performance and stable user behavior. About 77.46% (SRP + TRP) of the user behavior was efficient. The task also showed a very good learning effect: Questionnaires before the task showed that all of the participants had no concept of the intramolecular forces between the atoms within the CO_2 and CS_2 Molecules. After the task, however, 10 of the 12 participants could distinguish the intramolecular forces and describe an accurate conception of the force. This indicates that the haptic features (i.e. force feedback, gravity wells) within the haptic interface were able to aid students with visual impairments in developing accurate conceptions of intramolecular forces. User Preference Questionnaires indicated that users liked the spring model of the intramolecular force, indicating that the gravity well and force feedback haptic features could enhance user understanding of the intramolecular force between atoms. However, more research is needed to define the intensity of the force, as some participants commented that the force was too weak, while others said it was too strong. Likewise, more usability research should be conducted to determine the optimal haptic interface design elements to greater facilitate intramolecular force conception for users with visual impairments.

6 Conclusion

This study investigated the effect of haptic interface and the learning effect. The result showed that haptic interface can significantly affect user performance as well as the learning effect of users with visual impairments. Moreover, the visually impaired students enjoyed the hands-on experience very much. However, the study also found some haptic interface design limitations, which necessitate further research in order to improve haptic user interface usability. This study should provide invaluable empirical data and some insights to the future research of haptic user interface design.

References

1. Pleis, J.R., Lucas, J.W.: Provisional Report: Summary health statistics for U.S. adults: National Health Interview Survey. National Center for Health Statistics. Vital Health Stat. 10 (2009)
2. Braille Institute, http://www.brailleinstitute.org/facts_about_sight_loss#2
3. Patton, J.M., Braithwaite, R.: Special Education Certification/Recertification for Regular Educators. Journal of Special Education 24, 117 (1990)

4. White, R.T.: Implications of recent research on learning for curriculum and assessment. Journal of Curriculum Studies 24, 153–164 (1992)
5. Paik, S.-H., Cho, B.-K., Go, Y.-M.: Korean 4- to 11-year-old student conceptions of heat and temperature. Journal of Research in Science Teaching 44, 284–302 (2007)
6. Cohen, T., Ben-Zvi, R.: Improving student achievement in the topic of chemical energy by implementing new learning materials and strategies. International Journal of Science Education 14, 147–156 (1992)
7. Stapp, Y.: Facilitating the acquisition of science concepts in L2. TEFL Web Journal (2003)
8. Brooks, F.P., Ouh-Young, M., Batter, J.J., Kilpatrick, P.J.: Project GROPE-haptic displays for scientific visualization. Computer Graphics 24, 177–185 (1990)
9. Paulu, N., Martin, M.: Helping your child learn science. American Chemical Society, Washington, DC (1993)
10. Sauer, C.M., Hastings, W.A., Okamura, A.M.: Virtual environment for exploring atomic bonding. In: Proceedings of Eurohaptics, Munich, Germany (2004)
11. Monkman, G.J.: An electrorheological tactile display. Presence: Teleoperators and Virtual Environments 1 (1992)
12. McLinden, M.: Haptic exploratory strategies and children who are blind and have additional disabilities. Journal of Visual Impairment and Blindness 98, 99–115 (2004)
13. Yu, W., Brewster, S.: Evaluation of multimodal graphs for blind people. Universal Access in the Information Society 2, 105–124 (2003)
14. Brewster, S.A., Murray-Smith, R.: Haptic human-computer interaction. Springer, Glasgow (2001)
15. Minogue, J., Jones, M.G., Broadwell, B., Oppewall, T.: The impact of haptic augmentation on middle school students' conception of the animal cell. Virtual Reality 10, 293–305 (2006)

Using Grounded Theory and Text Mining to Find Interesting Reading Materials for Slow EFL Learners

Yuh-Chang Lin[1], Chia-Ling Hsu[2], Mu-Hua Lin[3], Hsiao-Fang Yang[3], and Chao-Fu Hong[4]

[1] Instructor, Center for General Education, Aletheia University, No. 32, Chen-Li Street, Tamsui, New Taipei City 25103, Taiwan, R.O.C.
[2] Graduate Institute of C&I, Tamkang University, No. 151, Ying-chuan Road, Tamsui, New Taipei City, 25137, Taiwan, R.O.C.
[3] Department of MIS, National Chengchi University, No. 64, Sec. 2, Chihnan Rd., Mucha Dist, Taipei 11605, Taiwan, R.O.C.
[4] Department of Information Management, Aletheia University, No. 32, Chen-Li Street, Tamsui, New Taipei City 25103, Taiwan, R.O.C.

au1258@mail.au.edu.tw, clhsu@mail.tku.edu.tw, {kikalin,hfyang.wang}gmail.com, au4076@au.edu.tw

Abstract. Many reasons contribute to slow EFL learners. Of all the reasons which cause slow EFL learners, reading materials not appealing to learners tend to be the one to blame. In order to help teachers to find out interesting reading materials for slow EFL learners in Taiwan, this study intends to use Grounded Theory [8] and Text Mining to search for the reading materials which are appealing to slow EFL learners in Taiwan. Based on the approaches of free voluntary reading proposed by Krashen [10], and extensive reading advocated by Day and Bamford [7], the theoretical framework of this study is established. Krashen claimed that when students read for information, pleasure or problem-solving, they have acquired the essence of free voluntary reading. Day and Bamford [7] argued that extensive reading is an important way to reinforce the concept of learning to read by reading. Both of these two reading approaches put emphasis on learners' free choice of reading materials which are within their linguistic proficiency, and at the same time, are interesting to them. Eighty-three university students are the participants, who took the second-year English reading course in a summer session program in Taiwan. They failed the course in regular semesters. This is their second- or third-time taking the same course. We call them slow learners accordingly. By the employment of Grounded Theory and Text Mining, it is expected that the interesting reading materials can be discovered and provided for teachers to adopt in their instruction in classes, and to a certain extent, learners' attention to reading passages can be drawn.

Keywords: Grounded Theory, Text Mining, interesting reading materials, EFL learners, free voluntary reading, extensive reading.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 77–85, 2011.

1 Introduction

1.1 The Importance of Reading

Reading plays an important role in literacy education and many research results have proven this. Krashen [10] advocated that his free voluntary reading is a way to develop one's literacy as well as to read for information, pleasure, and general knowledge. Day and Bamford [7] proposed that extensive reading is an approach to literacy education and only by extensive can learners acquire knowledge and information subconsciously or even unconsciously. Both of these two reading approaches share a common ground: Readers can choose what they like to read and read abundantly. For first language literacy education, free voluntary reading and extensive reading are widely employed to achieve the goal. However, it is not the case for reading in EFL contexts. This study aims to find out interesting reading materials for EFL learners with a view to attracting their attention to reading. Eventually, they may read extensively and at the same time they may develop their linguistic competence as well as their knowledge of the world.

1.2 Current Status of English Education in Taiwan

Currently, most studies on EFL reading in Taiwan, especially in the stage of primary and secondary English education, tend to focus on how to improve students' linguistic competence. It is this purpose that many teachers as well as materials designers make their efforts to use or produce materials that focus simply on the form of a language rather than the meaning of it. In other words, much emphasis is put on the grammatical aspects of a language: recognizing parts of speech for a word, distinguishing SVO grammatical pattern (subject-verb-object sequence of an English sentence), and acquiring phonetic system of English. Most of the efforts made both by teachers, materials designers, and students are for the purpose of mastering the form of English. Meaning of the content is rarely emphasized. That is the reason why students cannot recall accurately what they have read when they are asked to talk about the content of the reading passages they have just read. Students did not expose themselves to reading sufficiently. What with the reading materials chosen by teachers are not within students' linguistic proficiency, and what with the reading materials selected by teachers are not appealing to students, lead to students' lack of reading, which impedes students' significant improvement in English language ability.

1.3 The Significance of Meaning

Acquiring meaning of a reading passage will gradually build up a person's cognition process. From the accumulation of one's cognition of the reading materials or any events happening around, one may establish his or her schemata by turning facts into knowledge. When a person has much knowledge of the world, plus required linguistic ability, he or she will read both accurately and fluently. Meaning here plays an important role in the formation of one's schemata or background knowledge. It is meaning rather than form of a language which is responsible for the acquisition of a language. Free voluntary reading and extensive reading are the efficient approaches to equipping learners with as much general knowledge as possible. However, how can we make these two approaches possible in a foreign language context is another issue? Going back to the shared ground for both free voluntary reading and extensive reading, we may find two principles: One is that readers can choose what they like to read, and the other is that readers read abundantly. This is the theoretical base for this study: Finding out the materials interesting to students in order that both free voluntary reading as well as extensive reading can be made possible. It is hoped that through this study, materials appealing to learners can be found and learners' attention can be drawn because they are interested in the materials, for the attention readers pay to particular elements of the text is associated with interest [12][13][14].

1.4 Situational and Individual Interest

Interest can be divided into two alternative characterizations: situational and individual or personal interest [1]. Situational interest is centered in the immediate environment and is typically regarded as transitory or fleeting [4]. Because situational interest entails getting learners' attention and keep them excited or enthused, it can be a positive influence in students' text-based learning. As to individual or personal interest, it is enduring and reaches into an individual's cognitive and affective nature [2][4]. Individuals' vocations and avocations are indicative of their personal interests [11]. Moreover, these deep-rooted interests are strongly associated with self-concepts and self-schemata [3]. It is the second type of interest that this study is searching for.

1.5 Grounded Theory

Grounded Theory comprises a systematic, inductive, and comparative approach for conducting inquiry for the purpose of constructing theory [5][6]. It can be divided roughly into three main stages: open coding, axial coding, and selective coding. In the stage of open coding, researchers are fracturing and analyzing the collected data. The core category or related concepts tend to emerge from the data collected in this stage. And then through theoretical sampling and selective coding of data and constant comparison of incidents or indicators in the data to elicit the properties and dimensions of each category, the theoretical saturation of the core and related concepts tends to be achieved. It is this constant comparing of incidents that an interchangeability of indicator can be obtained, which means that no new properties or dimensions are emerging from continued coding and comparison. At this point, since the theoretical situation of the concepts has been achieved, the researchers shift the attention to the emergent fit of potential codes that make the conceptual integration of the core and related concepts to produce hypotheses that help explain relationships between concepts, which accordingly accounts for the latent pattern of social behavior that form the basis of the emergent theory.

1.6 Text Mining

Text mining is a process to acquire high-quality information from text. High quality in text mining usually refers to some combination of relevance, novelty, and interestingness, which is typically derived through the divining of patterns and trends by means of statistical pattern learning. Text mining involves the process of structuring the input text, which refers to paring, along with the addition of certain acquired linguistic features and the removal of others, and subsequent insertion into a database. Typical text mining tasks include text categorization, text clustering, concept or entity extraction, sentiment analysis, document summarization, and entity relation modeling.

1.7 Slow EFL Learners

Participants are 83 university students who took second-year English reading course in the summer session. This is a make-up class for students who have failed in regular semesters. Most of them took this course for second time, and part of them even takes the course for more than once. This is the reason why we call them slow learners. Their linguistic competence is not good, nor is their general knowledge of the world. Because of the lack of their linguistic competence, they find it hard for them to read in English. Grabe [9] argued that readers will employ the schemata or background knowledge acquired in the first language to read in a second or foreign language when their "level of second language (L2) proficiency has been developed first so that first language (L1) academic language abilities can more readily transfer and have an impact on L2 reading (pp. 145-146)". Interest and motivation are two other factors which will influence readers' willingness to reading. If readers are interested in certain topics, they tend to read them regardless of their poor language ability, for interest helps ease their difficulty with the language and makes extensive reading possible. In other words, with interest, readers may read for pleasure, information, and solving problems [10].

2 Procedure

Eighty three students are asked to write out their idea on two questions: Do you like to read? If you do, what do you like to read? Students answered these two questions in Chinese for around three paragraphs within an hour. After they have finished writing up their answers to these two questions, the data were collected and typed accordingly. Then open coding was conducted. In this stage, the researcher was fracturing and analyzing the collected data. First of all, terms related to topics and genres which appeal to learners are kept and conceived of as meaningful concepts. For example, terms such as novels, magazines and comic books are kept and taken as subcategories of genres. Additionally, reasons for the preference for certain types of genres or topics are also kept in order that new, interesting and novel scenario can be discovered. After this process, axial coding has been administered to find out new, interesting, and novel storylines which are embedded in the data collected. Finally, selective coding was conducted to find a scenario which is interesting and novel for the whole data collected.

3 Findings

By means of the application of human machine interaction, via three-stage coding process derived from the grounded theory, and through the employment of VFT Textmining System, we obtained four figures. They are shown as follows:

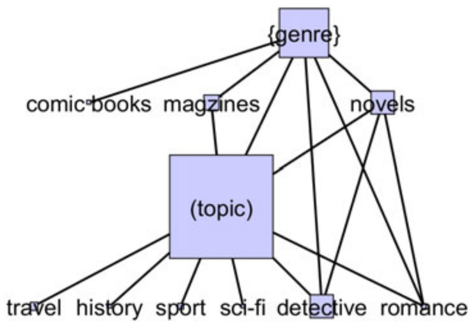


Fig. 1. A whole view of the extracted data

From the first figure, we may clearly observe that novels, magazines, and comic books are the main three categories which appeal to slow EFL university learners. These three categories can be referred to as the subcategories of genres. In other words, genre is the superordinate of novels, magazines and comic books. Also, we may see that novels have two subcategories: romance and detective. As to the subcategories of science fiction, sport, history, and travel, we are not sure whether they belong to the categories of novels or magazines. We cannot figure them out until we go on to observe the second figure.

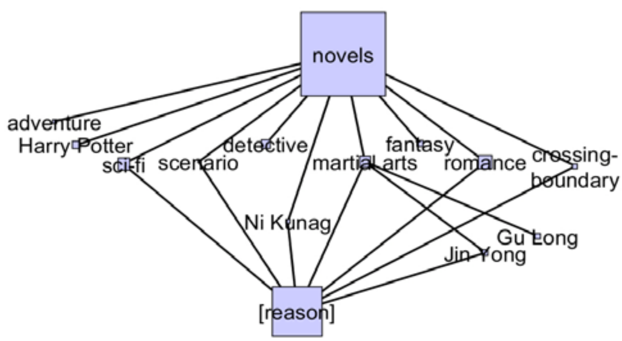


Fig. 2. Subcategories of novels

From the second figure, we may see that novels include such subcategories as Harry Potter, detectives, adventures, science fictions, fantasies, martial arts, romance, and crossing boundary. From this figure, we may find that science fiction belongs to the category of novels, and the rest of three subcategories (travel, history, sport) found in the first figure remains unidentified. Before we move on to find out the result, we want to point out two interesting discoveries. One is that students like to read Harry Potter; the other is that such Chinese authors of martial arts as Gu Long and Jin Yong are popular with students. Another Chinese author of science fiction, Ni Kuang was also admired by students. There is another finding that some students like to read so-called 'crossing-boundary novels', which is actually a kind of Internet novel mixed with the elements of martial arts, romance and fantasy. By utilizing this figure, we may come to a conclusion that the subcategories of science fiction and history are connected to the category of novels. Obviously, the category of magazine covers the subcategories of sport and travel.

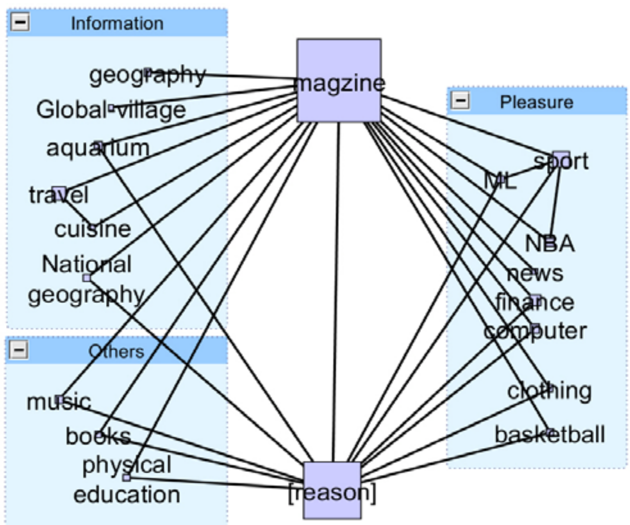


Fig. 3. Subcategories of magazines

In the third figure, we may find that below the category magazine, there are various kinds of subcategories such as newspaper, finance, computer, clothing, aquarium, basketball, Global Village (a kind of language learning magazine in Taiwan), travel, cuisine, National Geography, National Basketball Association, sport, physical education, music, and Chinese Professional Basketball. Of them, sport and travel are two main subcategories. We can roughly divide these subcategories into three items: information, pleasure, and others. Information covers items such as geography, global village, aquarium, travel, cuisine, National Geography. The items of pleasure include sport, major league, National Basketball Association, news, finance, computer, clothing, and basketball. Others contain such items as music, books, physical education.

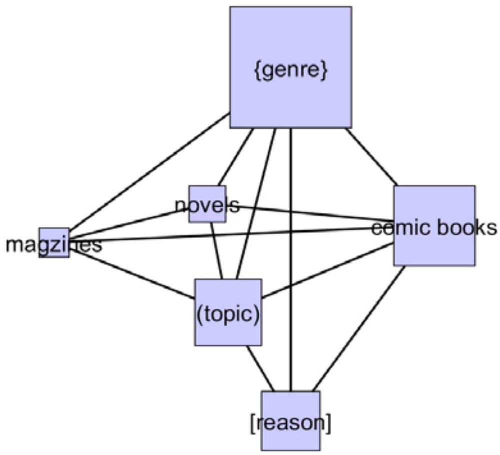


Fig. 4. An association map built based on the main category of comic books

There is no subcategory derived from the treated data, since students just mentioned comic books and the subcategories were not further elicited.

4 Discussion

As far as genre is concerned, novels, magazines, and comic books are the three main categories which appeal to students. When we take close look at these categories, we may find that in the category of novels, students like to read something about romance, detective, sci-fi, and history. In the category of magazines, students prefer sport and travel. For students who like to read novels, we may find a phenomenon that they like to read them both in traditional written form (namely books) and in the virtual world. Harry Potter and martial arts series are the ones in the traditional written form, and the so-called "boundary-crossing novels' are the combination of martial arts, romance, and fantasy, which are shown in the virtual world, that is, on the Internet world. This is a reflection of the computer generation, in which virtual worlds are created for people to fulfill some of their dreams which they fail to achieve in the real world.

Of all the subcategories connected to the main three categories (novels, magazines, and comic books), crossing boundary is the one which worth noting for the following two reasons. Firstly, it is at least new to me, even though I have been teaching English for almost 23 years. Secondly, crossing-boundary is a byproduct of the computer generation for it is a combination of romance, martial arts, and fantasies, which is presented in the virtual world. We may refer this phenomenon to the easy accessibility to computers. People increasingly count on computers for the management of their daily routines. They may read on the Internet for a variety of materials. Crossing-boundary novels in the virtual world are the reading materials appealing to university students. As teachers, it is one of their jobs to understand their students' interest so that they may provide their students with reading materials they are interested in.

From both categories and subcategories derived from this study, we may come to a conclusion that learners are interested in reading for either information or pleasure. For example, in figure 3, we may find that the subcategories of magazines can be roughly divided into three items: information (including geography, Global Village, aquarium, travel, cuisine, and National Geography), pleasure (including sport, ML, NBA, news, finance, computer, clothing, basketball, and others (including music, books, physical education). In the first sight, learners just read for information and pleasure, how can they improve their language ability? The answer is definitely confirmed by Krashen [10], for he argues that reading makes learners not only acquire the general knowledge of the world but also learn the language subconsciously.

Furthermore, from the following frequency table which is derived from the third figure, we may see the frequency of the items under the information category. The frequency for travel is 9, cuisine 5, aquarium 4, national geography 3, global village 2, and geography 2. When choosing teaching materials for learners to read, we should take the frequency of the items into consideration. Especially, when all the topics cannot be adopted in a time, we should cover the one or ones whose frequency is or are higher than others.

Table 1. Frequency table of the term under the category of information

Term	Frequency
travel	9
cuisine	5
aquarium	4
national geography	3
global village	2
geography	2

5 Limitation of the Study

The data collected in this study was limited, for the 83 students were from a class in a university in Taiwan. Even though they were slow EFL learners as stated above, they were actually not the whole population of a kind in Taiwan. However, the data analyzed are still worth paying attention to, for they stand at least for certain portion of learners who express themselves directly concerning their preference for the topics and genres when they read. Of course, more extensive and multi-dimensional sampling is needed in order that the theoretical saturation can be achieved.

6 Further Study

Finding out the interesting topics or genres for those slow learners is the first step to get slow EFL learners involved in the learning of a second language. However, it is not sufficient for language teachers to simply provide their students with the materials which appeal to their students. As is the case when teachers provide students with the topics or genres which students are interested in, students still find themselves unwilling to or incapable of reading the materials provided by the teachers because of too many unfamiliar words, phrases or sentence patterns they encounter while they are reading. It is a further study for researchers to search for materials which are within students' language ability, especially within students' word power. When learners read what they like to read and read without frequently stopping to look up unfamiliar words, they may read fluently. When they read fluently, their interest in the activities of reading may be stimulated. When students are interested in reading, they may read abundantly. When students read abundantly, the essence of extensive reading advocated by Bamford, Day and Krashen can be achieved. When students read extensively, the ultimate goal of reading for pleasure, information, and general knowledge [10] can be made possible.

References

1. Alexander, P.A., Jetton, T., Kulikowich, J.M.: The interrelationship of knowledge, interest, and recall: Assedding a model of domain learning. Journal of Educational Psychology 87(4), 559–575 (1995)
2. Alexander, P.A., Kulikowich, J.M., Schulze, S.K.: How subject-matter knowledge affects recall and interest on the comprehension of scientific exposition. American Educational Research Journal 31, 313–337 (1994)
3. Alexander, P.A.: Knowledge seeking and self-schema: A case for the motivational dimensions of exposition. Educational Psychologist 32, 83–94 (1997a)
4. Alexander, P.A.: Mapping the multidimentional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In: Pintrich, P.R., Maehr, M.L. (eds.) Advances in motivation and achievement, vol. 10, pp. 150–213. JAI Press, Greenwich (1997b)
5. Charmaz, K., Henwood, K.: Grounded theory in psychology. In: Willing, C., Station-Rogers, W. (eds.) Handbook of Qualitative Research in Psychology. SAGE, London (2007)
6. Charmaz, K.: Constructing Grounded Theory: A practical guide through qualitative analysis. SAGE, London (2006)
7. Day, R., Bamford, J.: Extensive Reading in the Second Language Classroom. Cambridge University Press, New York (1988)
8. Glaser, B.G., Strauss, A.L.: The discovery of grounded theory: Strategies for Qualitative Research, Chicago, Aldine (1967)
9. Grabe, W.: Reading in a Second Language: Moving from Theory to Practice. Cambridge University Press, New York (2009)
10. Krashen, S.D.: The Power of Reading, 2nd edn. Heinemann, Portsmouth (2004)
11. Mitchell, M.: Situational interest: Its multifaceted structure in the secondary school mathematics classroom. Journal of Educational Psychology 85, 424–436 (1993)
12. Reynolds, R.E., Shirey, L.L.: The role of attention in studying and learning. In: Weinstein, C.E., Goez, E.T., Alexander, P.A. (eds.) Learning and Study Strategies: Issues in Assessment, Instruction, and Evaluation, pp. 77–100. Academic Press, San Diego (1988)
13. Reynolds, R.E.: Learning important information from text: The role of selective attention. Review of Educational Psychology 4, 345–391 (1992)
14. Shirey, L.L.: Importance, interest, and selective attention. In: Renninger, K.A., Hidi, S., Krapp, A. (eds.) The role of interest in learning and development, pp. 281–296. Lawrence Erlbaum Associates, Killsdale (1992)

CAI Platform for Fundamental Geometric Training on Perspective Sketching

Ding-Bang Luh and Shao-Nung Chen

Department of Industrial Design, National Cheng Kung University
No.1, University Road, Tainan City 701, Taiwan, R.O.C.
luhdb@mail.ncku.edu.tw,
nomoloschen@gmail.com

Abstract. For most designers, freehand sketching is the primary tool for conceptualization in the early stage of the design process. However, existing education programs on concept presentation techniques rarely emphasizes the practicing of two most fundamental geometric shapes - cube and cylinder. Lack of correct reference and proper training, students often end up with disproportioned sketches that deviate from common visual experience which would lead to misunderstanding of original design.This research developed a computer-assisted cube sketching instruction platform for novice, with which users can practice correct cube sketching using freehand skills or digital devices for self-improvement. This platform can provide instant corrective feedback and demonstrated 19% sketch accuracy increased relative to the control group in a series of experiments. Based on the successful experience, a cylinder training program is under development. The ultimate goal of which is to develop a comprehensive CAI platform to help novice improve their skills by self learning and correction.

Keywords: CAI, Perspective sketching, Self-instruction, Geometric sketch practice.

1 Introduction

Freehand sketching has long been recognized as an indispensible element in product design and development. Sketches can provide the designers with visual clues that inspire creative inventions, generate new information, and expand mental imagery(Goldschmidt, 1991, 1994). Akin in his 1978 article pointed out that sketches can help designers adjust and synthesize ideas in the problems solving process. In fact, sketches facilitate the visual search among alternative options and the exploration of design concepts.

In a common instructional curriculum for sketching, students often starts with simple but fundamental geometric units before developing the sketching skills for complex shapes (Henton, 1980). Each and every complex shape can be treated as dissected or stacked cubes (Liu, 1997). Therefore, the cube is the most fundamental shape in design. Errors of perspective cube sketching can be broken into 7 types: 1) askew vertical line, 2) anti-perspective, 3) askew horizontal line, 4) beyond cone of

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 86–92, 2011.

vision, 5) excessive vanishing points, 6) proportion maladjustment, and 7) vanishing lines in parallel (Luh and Yang, 2002).

However, the lack of an accurate mechanism for inspecting student sketches and providing a geometrically correct solution as reference means that even instructors and professional designers can only correct and offer suggestions to their students based on their experiences. Discrepancies between such feedbacks and truly accurate solutions cannot be readily determined unless the sketches are inspected individually using the perspective principle.

Most existing sketch related CAI systems research focus on fundamental knowledge instruction including basic perspective, two point perspective, light-shadow perspective, and geometry perspective(Wu, 1998； Hong, 2000； Lin, 2004； Lin, 2004). These systems often concentrate on repetitive and fixed materials so the students can study the knowledge-based subjects after school, and find textbook solutions in standardized databases. Such CAI systems provide learning opportunities unconstrained by time or space, and can therefore provide substantial educational assistance where resources are severely lacking.

Motivated by the fundamental relationship between freehand sketching and perspective as well as CAI systems' potential to achieve active learning and compensate for limitations in educational resources, this research identifies the specific needs of sketch learning and incorporates them into a software instruction system. The resulting product utilizes two point perspective as the fundamental principal, focuses on the cubic shape as its subject, and demonstrates the ability to inspect and offer revision suggestions to users' sketches, thus providing real-time error detection and customized feedback to assist users in self-directed active learning.

2 Reversed Perspective Approach

The basic components of foot-point approach two-point perspective drawing includes (Figure 1): from the top view, the location of picture line (PL) (to simplify, PL in the

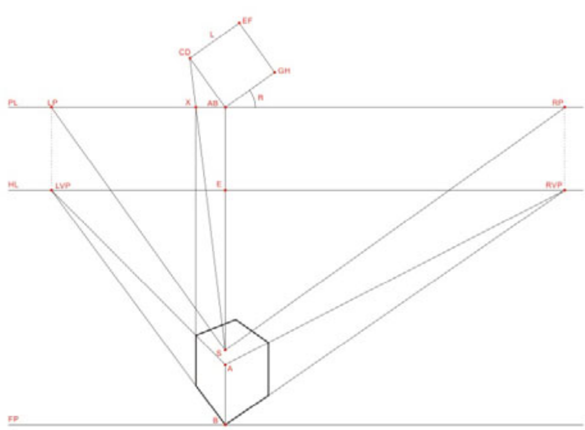


Fig. 1. The foot-point approach two-point perspective

figure is aligned with the front edge of the unit to maintain its actual length), unit length (L), rotation angle (R), station point (S), and the length between PL and the viewer (segment length of SA); from the side view, the locations of the horizontal line (HL) and foot-point (FP), and the distance between the object and the viewer's eye (E) or foot-point (in this case the unit is placed on the ground).

In accordance with the above drawing information, reversed drawing process (Figure 2) is listed as follows:

1. Extend perspective line to obtain vanishing point (LVP). From point LVP, draw horizontal line HL perpendicular to the front vertical line AB; extending line bh results in the vanishing point on the other side (RVP).
2. The cone of vision can be defined by drawing a circle with a radius of half the length between two vanishing points. The station point (S) can be identified by extending the front vertical line AB to the circle of the cone of vision.
3. If there exists a picture line (PL) which is parallel to the horizontal line HL, the front vertical line AB extension intersects with the picture line PL at point X. Draw two line segments XY and XW, with lengths identical to that of line AB, parallel to line SRVP and line SLVP respectively.
4. Extend line CD to intersect with PL at point J; draw line WJ to intersect line AB at the standing point (SP)
5. Line SPY intersects with picture line PL at point K. The vertical line of point K intersects with line BH at point H', which is the correct location of point H. Follow the same method to define the correct location of point G.
6. The intersection of line G'RVP and line CLVP is the correct location of point E, while the intersection of line DRVP and line H'LVP is the correct location of point F.

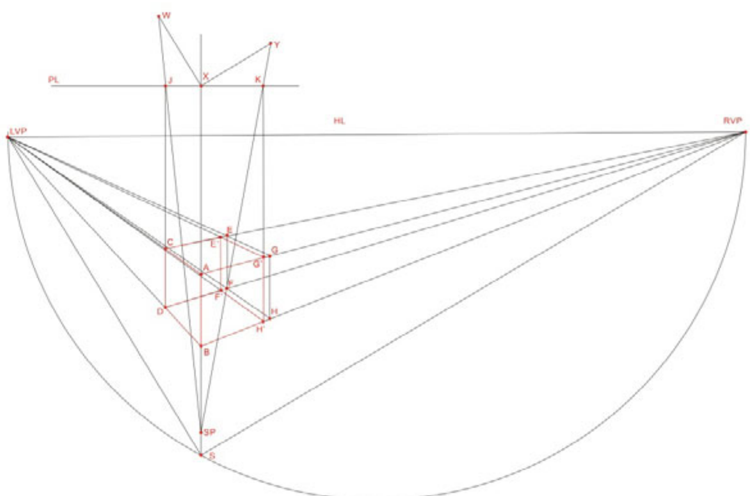


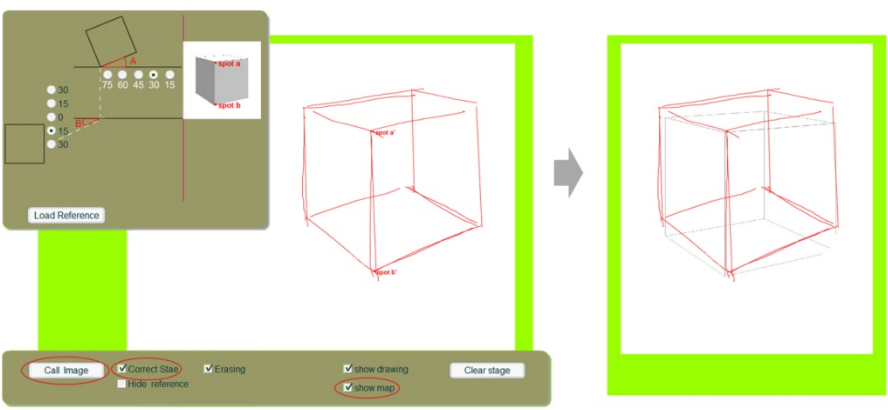
Fig. 2. The reversed two-point perspective approach

3 System Components and Interface

3.1 Introductory Training Program

The development principle for this program is “Observe and memorize.” By providing a verity of reference images for the user to observe and sketch, this program develops the user’s observation, hand-eye coordination, and visualization ability for cube sketching. Based on the cone of vision, the program offers five horizontal rotation angles (15, 30, 45, 60, and 75) and five vertical rotation angles (30, 15, 0, -15, and -30), for a total of 25 three-dimensional views of an orthogonal cube. This helps the user develop accurate spatial understanding of a cube.

After starting the introductory training program, the user can choose the cube he or she wants to practice with, specified by the horizontal rotation angle A and vertical rotation angle B, from the upper left-hand corner of the reference image panel. Clicking on “Load Reference” would bring up the chosen cube to the right-hand field. The user can also hide the reference image by checking “Hide Reference”; this can test the user’s memory, visualization, and understanding of the cube. Based on personal preference, the user can choose the input method (digital pen or conventional pen) to sketch the chosen cube in the specified field. After sketching the cube, the user can compare the sketch with the correct solution by checking “Correct State” and clicking on the reference vertices of the sketch and the reference image (i.e., points A and B and points a and b in Figure 3(a)). If the user clicks on “Call Image” when “Show Map” is checked, the correct solution and the freehand sketch would then superimpose on one another for comparison and correction (Figure 3(b)). After making corrections, the user can then click on “Clear Stage” to clear the image for the next training lesson.



(a) (b)

Fig. 3. The operating interface of the introductory training program: (a) input the reference vertices of the sketch and the reference image; and (b) the superimposition of the correct solution and the freehand sketch

3.2 Advanced Revision Program

The development principle for this program is “Visualize and Verify.” The user can specify the perspective angle of the cube, let the system detect and point out errors, and construct accurate images as references. The objective is to enhance the user’s accuracy in sketching cubes from imagination.

After starting the Perspective Practice revision program (Figure 4, left), the user can first adjust the error tolerance based on his or her skill level or need. This can be done by adjusting X (askew angle of the vertical line), Y (tilt angle of the horizontal line), and Z (vanishing line error quotient). The user can then sketch a cube from imagination onto the drawing field, and check “Locating Spot” to mark the eight reference vertices (Figure 4, middle) before clicking on “Correcting.” The program would then identify errors with red lines and/or descriptions, and offer revision suggestions/instructions in the yellow “System Message” box on the right-hand corner of the screen. The user can also check “Hide Message” to practice his or her ability to identify errors without assistance. After inspecting the sketch, the user can click on “Clear Stage” to clear the image and decide whether to adjust any program setting. The process can be repeated until the user makes no error outside of the specified error tolerance.

Once the sketch passes error inspection, the user can click on “Re-standardize” to reveal the correct solution superimposed in blue lines (Figure 4, right) for detailed comparison and further modification. Finally, the “Calculate” function allows the user to calculate the area of the sketched cube, and the percentage of sketching error. The user can save the data for future reference or analysis by clicking on “Save Data.”

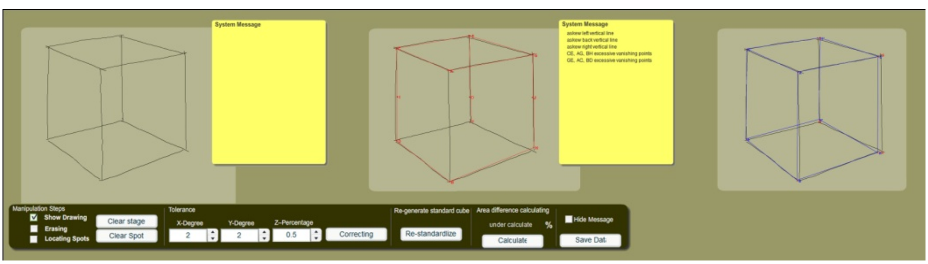


Fig. 4. The operating interface of the advanced revision program: Advanced program setting options; reference vertices marking and revision instructions; superimposition of the correct solution shown in blue

4 System Effectiveness

This study analyzed the Experimental Group’s drawing ability before and after the experiment. The percentages of error were summed up into cumulative percentage of error, and both cumulative percentage of error and percentage of accuracy pre- and post-experiment were compared through Paired-Samples T Test. As summarized in Table 1, both cumulative percentage of error and percentage of accuracy were significantly improved (Sig. < 0.05), indicating that the CAI system is effective in reducing students' errors in sketching cubes. Comparing the percentages of accuracy pre- and post-experiment (0.15 and 0.34, respectively), the study showed that this CAI system improved sketching accuracy by 19%.

Table 1. Experimental Group's cumulative percentage of error and percentage of accuracy before and after the experiment

		Average	t	Sig. (2-tailed)
Cumulative percentage of error	Pre-experiment	2.07	-3.336	0.004
	Post-experiment	1.17		
Percentage of accuracy	Pre-experiment	0.15	-2.579	0.020
	Post-experiment	0.34		

5 Conclusions and Suggestions

The current research develops a computer-assisted instruction (CAI) system based on foot-point approach two-point perspective. The system is capable of identifying sketching errors and providing revision suggestions, and the teaching experiment has arrived at the following three conclusions:

1. This research has developed a CAI system suitable for technical training; the system can provide correct reference solutions and revision suggestions, enabling effective learning through the system.
2. This system employed reversed perspective approach to derive technical drawing from perspective drawings, and subsequently reconstructs correct perspective reference solutions for users to compare and revise their sketches. Experimental results demonstrated the feasibility and effectiveness of this approach.
3. When the CAI system developed in this study was tested in a learning experiment, the Experimental Group improved sketching accuracy by 19% (2.35-fold higher than the pre-experiment accuracy), and demonstrated significant improvement compared to the Control Group.

This study is an initial attempt to develop a CAI sketching system and determined its feasibility, practicality, and effectiveness in technical education. As such, the system developed in this study focuses on the training of two-point perspective cube sketching. To advance this effort, the program can be expanded in the future to include cylindrical, conical, pyramidal, and other geometric shapes, making the training system more comprehensive, and effectively enhance the program's function and the users' learning experience.

References

1. Akin, O.: How do architects design?, Artificial Intelligence and Pattern Recognition in Computer Aided Design. North Holland Publishing Company, Amsterdam (1978)
2. Goldschmidt, G.: The dialectics of sketching. Creativity Research Journal 4, 123–143 (1991)

3. Goldschmidt, G.: On visual design thinking: the vis kids of architecture. Design Studies 15, 158–174 (1994)
4. Henton, R.W.: Quick sketch: a new technique in interior design graphics. Kendall/Hunt, Dubuque (1980)
5. Hong, W.S.: The research of design and effectiveness of computer assisted instruction for perspective. Unpublished master's Thesis, Tatung University, Taipei, Taiwan (2000)
6. Lin, C.H.: The effectiveness and design of computer-assisted instruction for descriptive geometry. Unpublished master's Thesis, Tatung University, Taipei, Taiwan (2004)
7. Lin, S.T.: The design and research of computer assisted instruction for light-shadow perspective drawing. Unpublished master's Thesis, Tatung University, Taipei, Taiwan (2004)
8. Liu, Y.S.: Presentation techniques for design. Liuho, Taipei (1997)
9. Luh, D.B., Yang, T.L.: Presentation techniques for creative ideas. OpenTech, Taipei (2002)
10. Wu, C.T.: The research and design of computer assisted instruction. Unpublished master's Thesis, Tatung University, Taipei, Taiwan (1998)

A Reading History Logger for Supporting Reading Habit Development

Yasuo Miyoshi[1] and Takaaki Oobayashi[2]

[1] Faculty of Science, Kochi University, Japan
[2] Graduate School of Integrated Arts and Sciences, Kochi University, Japan
{miyoshi,obayashi}@is.kochi-u.ac.jp

Abstract. We are developing a social networking site for the purpose of support for a reading habit development. To support the habit development, it is necessary to grasp the state of the user's reading in real time. However, letting a user register every reading history is a big burden for the user aiming at the habit development. Therefore we have developed a bookmark-style reading history logger device using Eye-Fi and PSoC. The bookmark device has a function to register the reading date/time information to any photo-sharing site. Our social networking site for a habit development can get the reading time by checking a web feed of the photo-sharing site.

Keywords: Habit Development, Reading, Logger, Peer Mentoring.

1 Introduction

Development of learning habit and improvement of learning motivation are important in self-directed learning activity. Activity of reading books is a one of self-directed learning activity. We focus peer mentoring[1] utilizing interaction-centered model in motivational design[2], and intend to design an environment for activities of peer mentoring. We aim to support development of learning habit with peer mentoring, and are developing an environment for peer mentoring on a social networking site. Gathering a state of the users is required to peer mentoring. However, it is nonsense that the user who wants to develop the reading habit records the history at every reading. Therefore our study considers a method to get the reading histories in some electronic book readers (e.g. Kindle, iPad) and traditional paper media books [3][4].

2 Reading Habit Development with Peer Mentoring

We have focused on a peer mentoring method which habit development applicants make mentoring with each other. Figure 1 shows peer mentoring. The applicant must observe a partner and grasp the state of the partner to comment to the partner in order to perform mentoring. However, it is impossible that the applicant always observe the state of reading of the partner in the case that the object of habit development is reading. They cannot give them appropriate advices with each other. Therefore we

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 93–96, 2011.

have proposed a peer mentoring system on a social networking site as shown in figure 2. The habit development applicant records one's reading history in the system as log instead of observing the reading state of the partner and record. The system analyzes the log and measures the appropriate timing for advising. If the appropriate timing comes, the system shows the reading state of the partner and gets the applicant advises the partner. We know that when learners are conscious of the existence of their familiar person, they can maintain their motivation in early learning activity[5]. The system does not show some automatic generated sentences as advice to the applicant because the repetition of generated advices without warmth negatively influence their motivations.

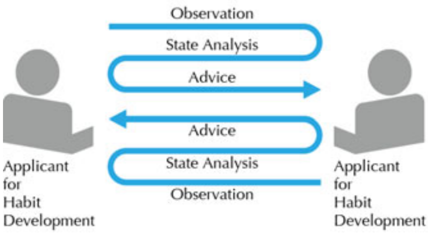


Fig. 1. Peer Mentoring

Fig. 2. Peer Mentoring System

3 Reading History

3.1 Reading History for Supporting Reading Habit Development

Ideally, we should get the reading history which has Five W's and One H. Example: {Who: Ichiro Tanaka. When: one hour from 19:00 on 7th July 2010. What: "Introduction to Statistics" (ISBN: xxxxxx), pages 15 to 80. Where: the library of xxxx University. Why: to write a report for the lecture of statistics. How: while taking the memo.} For supporting reading habit development, the required information should be for confirming whether or not the user read a book. That is we must get the user information (who) and time information (when). The requirements for getting time information are able to get start (restart) and finish (abort) time automatically and promptly.

3.2 To Get a Reading History

We consider to get reading history in paper media books, because electronic book readers are unfamiliar and there are few digitalize book data. It is difficult to get a reading history except time information from paper media. We propose a bookmark-style reading history logging device for getting automatically the start time and the finish time. This device is used like an original bookmark and held between the pages of a book. It detects opening and shutting of the book and records the time as start time and finish time of reading. Therefore it can record automatically the time of reading without user's operation. Furthermore it can get the information promptly if it uses a storage that has a wireless communication function.

4 Prototyping a Bookmark-Style Reading History Logging Device

We have prototyped a bookmark-style reading history logging device for getting a reading history on a paper media book. It detects opening and shutting of the book by an optical sensor (CdS cell). The end of the sensor connects with tied optical fiber because the sensor has thickness. As shown in figure 3, a bundle of optical fiber is laid slimly and closely together. This thin part is tucked into the book. This bookmark device does not detect a light while into the shut book, but can detect it while opening the book. At this time, the microcontroller writes the time, as the reading time, onto Eye-Fi SD Card. The microcontroller embeds the reading time, as Exif metadata format, in a dummy picture file, because the Eye-Fi has a function to upload

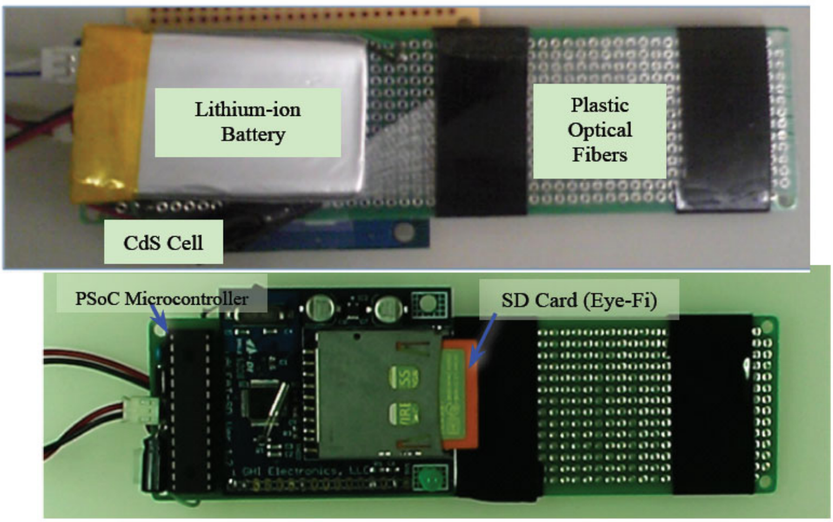


Fig. 3. Reading History Logging Device (front side and back side)

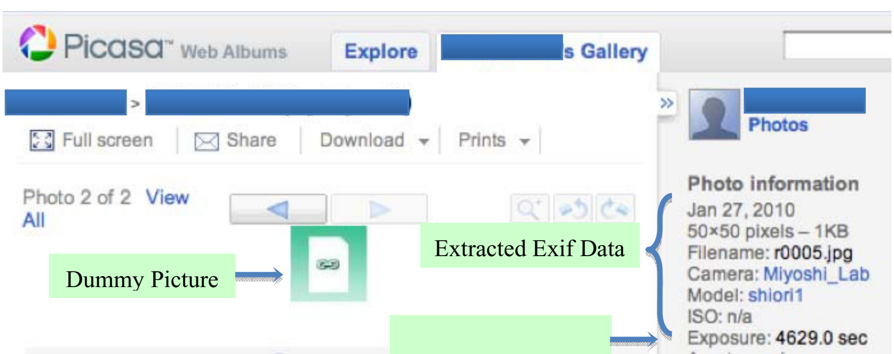


Fig. 4. Extracted Exif Data from a Dummy Picture on a Photo-Sharing Site

wirelessly and automatically photo files. When the dummy picture including the reading time is written in the Eye-Fi, in the enabled wireless LAN area, the Eye-Fi card uploads the picture to any supported photo-sharing website automatically. The photo-sharing site that received the dummy picture should probably be able to extract the Exif data from the dummy picture (figure 4). It can be got the reading time information that is included in Exif data by checking a web feed of the photo-sharing site.

5 Summary

Although our prototyping device can only record user's reading time, it is used like an original bookmark, and can record automatically the time of reading without user's operation. We have confirmed that our social networking site that is for supporting reading habit development can get automatically users' reading histories that were recorded by the prototyping device. We think that data logging using Eye-Fi like our method is useful for other data logging like a lifelog.

References

1. Baker, W., McNicoll, A.: I can get by with a little help from my friends: Peer mentoring - critical friends for the reflective practitioner. In: Proc. of the 2006 Ann. Intr. Conf. of HERDSA, pp. 28–34 (2006)
2. Keller, J.M.: Motivational design. In: Unuwin, C. (ed.) Encyclopaedia of Educational Media Communications and Technology, 2nd edn., pp. 406–409. Greenwood Press, Westport (1988)
3. Amazon.com, Inc.: Kindle Development Kit for Active Content, `https://kindlepublishingamazon.com/gp/vendor/members/kindlepubs/kdk/home`
4. Apple Inc.: Start Developing iPad Apps, `http://developer.apple.com/ipad/sdk/`
5. Miyoshi, Y., Okamoto, R., Kanenishi, K., Yano, Y.: A Design of Social Networking Service for Supporting Learning Habits Development. In: Proc. of IADIS International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2009), pp. 396–399 (2009)

A Drawing Learning Support System with Auto-evaluating Function Based on the Drawing Process Model

Takashi Nagai, Mizue Kayama, and Kazunori Itoh

Graduate School of Science and Technology, Shinshu University
4-17-1 Wakasato, Nagano 380-8553 Japan
nagai@ngi644.net, {kayama,itoh}@cs.shinshu-u.ac.jp

Abstract. The purpose of this study is to develop a drawing learning support system using a networked environment. In this paper, first, we show the outline of the online drawing learning support system. Second, we describe the drawing process model that support individual drawing learning. Finally, we show three examples of learning with our system.

Keywords: Drawing , Learning Support System, Drawing Process Model.

1 Introduction

Art education in a networked environment has been introduced recently. However, there are some limitations in the functions and content of tools for basic skill learning such as drawing, painting, and sculpturing [1]-[3]. Drawing is one of the fundamental skills in art education. All beginners must acquire these kinds of skills first [4],. Learning related to art requires repeated practice with a trial-and-error process [5],[6]. Therefore, to learn drawing is categorized as skill-learning [7]. In this type of learning, novices cannot recognize whether or not they draw correctly and appropriately.

The purpose of this study is to explore a support system for beginners in drawing. In this paper, we show learning flows in our system, then we describe some examples of learning with our system.

2 Online Drawing Learning Support

2.1 Learning Flows

In this study, the learner's drawing process that is recorded by a digital pen is reused in order to replay learner's drawing process [8]. A learning activity is started after the tutor defines a learning task in the learning management system (LMS). The following flows are ideal learning processes in our learning environment shown in Fig. 1:

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 97–106, 2011.

1. A learner draws his/her work.
2. Both of the learner's drawing process data and his/her work are registered in the LMS.

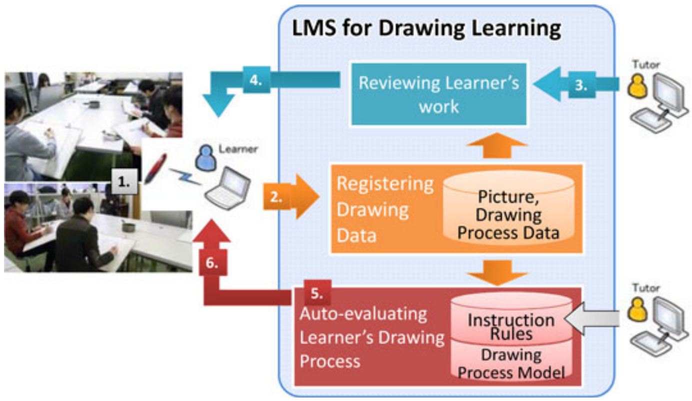


Fig. 1. Learning flows with the proposed system

[Interaction with Learner and Tutor via System]

3. A tutor evaluates the learner's work by replaying the learner's drawing process, then the tutor gives some advice on both the learner's drawing process and his/her work.
4. The learner takes the tutor's comments about his/her own drawing process and work.

[Interaction with Learner and System]

5. The LMS diagnoses the learner's drawing process based on the drawing process model (DPM) and the instructional rules that are defined by tutors.
6. The learner is given an auto-evaluated result from the diagnosis.

2.2 The Drawing Process Model

Our system arranges the learners' drawing processes by using the drawing process model (DPM). The DPM is developed by an inference engine that is able to detect three drawing phases. This model consists of 3 types of parameters. They are the drawing step, the drawing phase and the features of the drawing strokes. Fig. 2 shows an outline of the DPM.

Seven Drawing Steps. In an interview with five art experts, we collected the drawing processes of experts. Then, we formulated the seven step model as a hypothesis for simplification of an artist's drawing process. The contents of each step are shown in the lower part of Fig.2.

"Drawing is seeing"[9]. Hence, the first step is to carefully observe the drawing subject. The relationship between the light source and the drawing objects is also checked in this step. In step 2, the relative locations of the objects are confirmed based on the vanishing points. In the next step, the composition of this picture is defined. The drawing area is fixed on the drawing paper. In step 4, the outlines of the drawing subject are expressed in simple lines in a balanced way. The rectangle, the oval, triangle, straight line, and simple curve are used consciously in this step. The size, the location, and the direction of each object are also pictured in this step. In step 5, shading is added. Various values of light and dark are expressed in the drawing. The shading techniques become complex for a square pillar, a cylinder, and a sphere in this order. Shading should be added first to objects whose outline shapes are square pillars, second to cylinder shapes and then finally to spheres. In step 6, a learner checks the material of each object, and then expresses its texture in the drawing. Finally, the finishing touches are added. A learner draws details of each object. The balance of the total subject is also considered in this step.

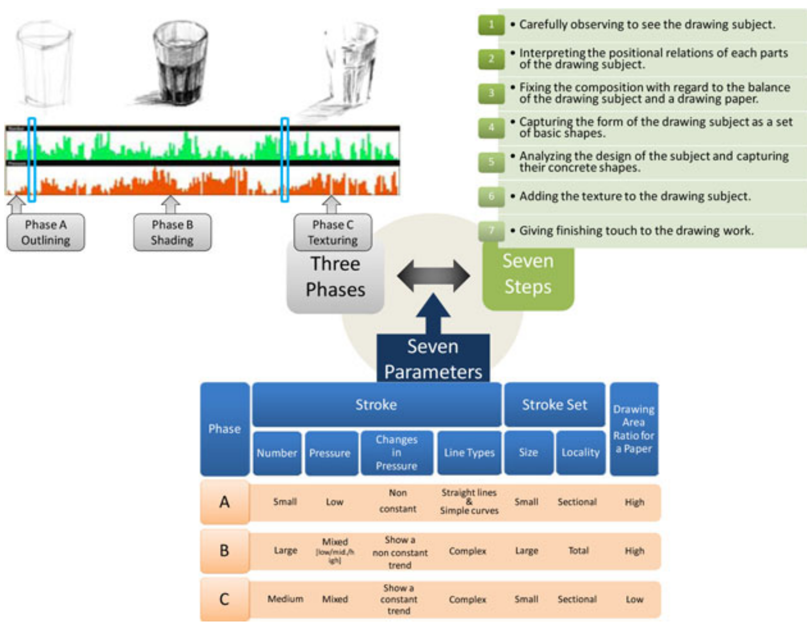


Fig. 2. The drawing process model

Three Drawing Phases. We found three different drawing phases from quantitative investigations of the time variation in the drawing strokes and the pen pressures of a digital pen used by art experts. The left side of Fig. 2 shows the features of strokes and pen-pressures of the drawing process. The features of each phase can be described as follows:

- Phase A [Outlining]: A learner interprets the drawing composition and the outline of the whole object by using simple lines to define the proportion in a perspective way.
- Phase B [Shading]: A learner draws the drawing subject totally and adds shade for whole parts.

- Phase C [Texturing]: A learner adds the texture of the subject in detail. He/she fixes the balance of the subject on the drawing paper.

The boundary of each phase in the drawing process is determined mainly based on the changes in pressure. In this study, the state of the pressure changes in writing a stroke of a drawing is called "changes in pressure". In an ideal drawing process, the artist outlines the motif in the first stage of his/her drawing (see Step 3 in drawing steps). At this stage, an ideal artist draws a rough line using low pen pressure, and draws the outline of motif using a high pressure line. Therefore, the pressure at this stage tends to be lower, and the state of change is not constant. So, the drawing section which satisfies the following three conditions is identified as Phase-A.

- There are multiple sections where the average pressure is zero or close to zero.
- The state of pressure change does not tend to be constant.
- The ratio of the area of the drawing in that section to the area of the final draw is more than 75%.

The process parts for Phase-B and Phase-C are selected from the whole drawing process other than the section which is assigned to Phase-A. At first, our system divides the selected parts into 10 blocks in the same time span. In each block, if its state of pressure change does not show a constant tend, the block is assigned to Phase-B. Also, if it shows a constant trend, the block is assigned to Phase-C. Then, the sequential blocks in the same phase are grouped. Two division lines that are detected by these rules are shown in the squared areas in the left part of Fig.2.

Seven Parameters for Drawing Stroke Features. We have to define concrete and objective features of each phase in order to define these three phases of the drawing process. The left part of Fig.2 shows the quantitative features of these phases. Seven parameters are shown in this table. They are the number of strokes, the stroke pressure, the changes in pressure, the line types, the degree of assembled stroke size, the dispersion of the drawing area, and the ratio of the drawing area. The number of strokes and the pen pressure are relative values in the drawing process. The size and the dispersion of the stroke sets are relative values on the drawing paper (or entire drawing area)

Each parameter is expressed in more than two levels. The number of strokes is expressed as small, medium, or large. The pen pressure is expressed as low, medium, or high. The changes in pressure are expressed as non constant or constant. The line types are point, straight-line, simple curve, or complex line (includes curve). The size of the stroke sets is large or small. The locality of the stroke sets is sectional or total. The ratio of the drawing area is high or low.

2.3 Drawing Evaluation with the DPM

Fig. 3 shows a drawing process viewer that our system provides. The upper part of this viewer is an area for replaying the drawing process and showing the instructional information. Two types of advice are added in this drawing. These are an instructional comment and an instructional drawing. The former is shown in the timeline bar in the viewer. A comment is linked to a specific point in time when the learner performs an inadequate drawing action. The latter is also connected to a point in time. In this case, an instructor adds red lines to point out error cpositions, and adds the comment "represent the outline of the vase". The lower part is a graph area. This area includes six graphs that indicate the seven features of the drawing process. They are the number of strokes, the pen pressure, the line types, the size of stroke sets, and the drawing locality and ratio of drawing area.

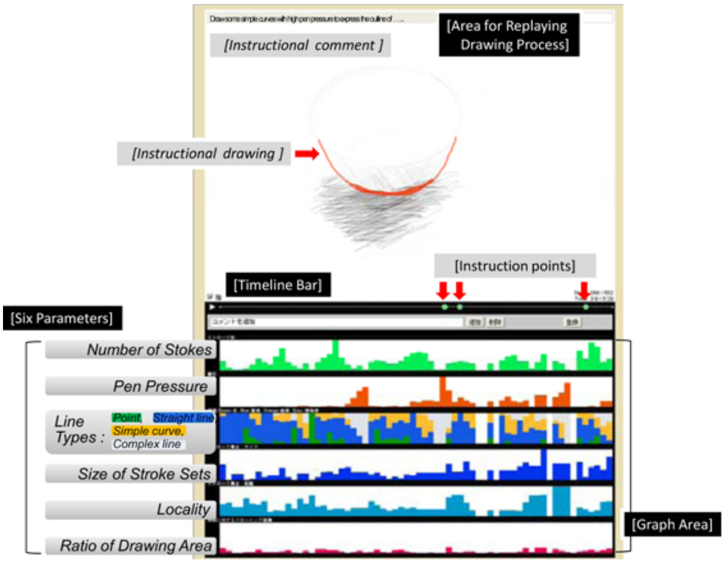


Fig. 3. An example of the drawing process viewer

The red arrow on the far left shows the first instructional comment for this drawing. Learners obtained the following advice from our system: "Draw some simple curves with high pen pressure to express the outline of the subject. Observe the drawing objects more carefully. Capture each object as a simple configuration."

3 DPM-Based Drawing Learning

3.1 Classroom Learning Support

Fig. 4 shows the drawing learning support functions of our system. Learners upload their drawing process data to our system after their drawing. The tutor of this class and all of the class members are able to access the portfolio pages for today's results (Fig. 4 upper left). A member of this class is able to refer to the drawing pictures and these process data in this portfolio. A drawing process viewer page is shown to him/her when a learner chooses a learner's drawing result (Fig. 4 upper right). A learner can replay the drawing process from anytime. This viewer is able to show graphical evaluation results for six parameters in time series graphs.

Our system generates some advice for the drawing process based on these results. The tutor's comments are also confirmed on this page. The evaluation results from both the system and the tutor are added to the time series. The learner can recognize which points are wrong and correct the drawing.

Referring to experts' and others' drawing processes helps a learner understand the advice from the system and the tutor. Moreover, comparing one's drawing process with others could help them find new techniques and/or drawing methods. Learners hand in their drawing processes and then obtain the tutors' feedback as individual learning support tools (Fig. 4 lower part).

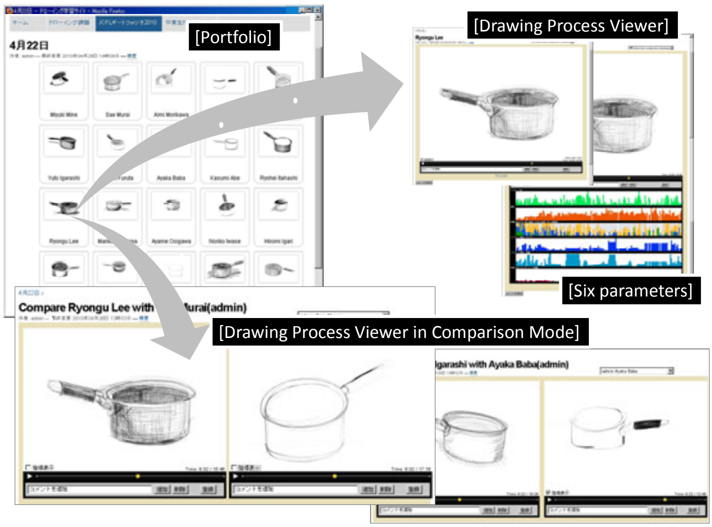


Fig. 4. The drawing learning support functions

3.2 Individual Learning Support

Overview. This function is used the between the time when a learner finishes his/her drawing and the time when a teacher gives this learner some direct instructions. To support the individual drawing learning, learners' drawing processes are analyzed based on the DPM. At first, our system separates the learner's drawing process into three drawing phases. Then, the geometrical features of each phase are compared with the DMP. Based on this analysis, our system generates advice for that learner. Timing and content of that advice are determined. There are two types of advice: instructional comments with short sentences and the instructional drawings as models.

Supporting functions. The supporting functions for individual drawing learning are as follows:

- Replaying one's own drawing process.
- Synchronous replaying of one's own drawing process with those of other learners / art experts / instructors.
- Showing the location of the 3 drawing phases in a learner's drawing process.
- Total advice comments for the whole drawing work.
- Instructional comments and drawing for a sectional drawing.

- Introduction to some drawing techniques and exercises which should be mastered by this learner.
- Introduction to the drawing work of other learners / art experts / instructors as a model.

Educational Effectiveness. The educational effectiveness of those functions was examined at an art school. The subjects were 18 students who had been at the school for 1 month. The knowledge and experience about drawing for each student was different. From the preliminary survey, we found that the 15 students were beginners. The subjects were asked to use our tool for a month for their individual learning. The frequency of use was not specifically mentioned.

In this experiment, we focus on the appearance of Phase-A in the learners' drawing processes. Phase-A is the most important part to sketch a motif. However, many instructors in art school do not tend to teach this matter explicitly to their learners. Before this experiment, Phase-A appeared in the drawing processes of 3 students who were not novices. The drawing processes of the other 15 students, by contrast, did not show Phase-A. The works in Fig. 5 are the final drawings of two students whose drawing process does not have Phase-A. Fig. 6 and Fig.7 show the drawing results in each phase of these students. The upper part of each figure shows a graph of the changes in pressure in a time-line and the detected phases. The lower part shows drawing results in each phase. In those results, two problems can be point out.

- Learners wrote some clear lines with high pressure from the beginning.
- Learners did not draw the whole shape of the motif at the beginning.

Thus, this learner was trying to show the form of a motif immediately.

During the experiment, our system checked for the appearance of Phase-A. If Phase-A did not appear, some text comments (for example "You should learn the. techniques so as to form a composition by drawing a simple shape.") are given to the learner at an appropriate time. The system also introduces a link to the related exercises. In these exercises, a learner is asked to draw some lines with the indicated pen pressure, and/or to draw the whole outline of the motif with low pressure. At the same time, the system suggests to refer to an expert's drawing process to form the composition in simple shapes.

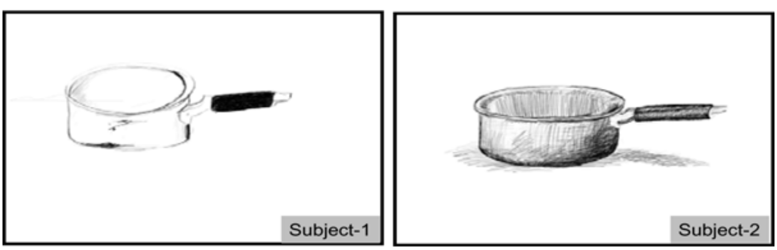


Fig. 5. The final works prior to the experiment

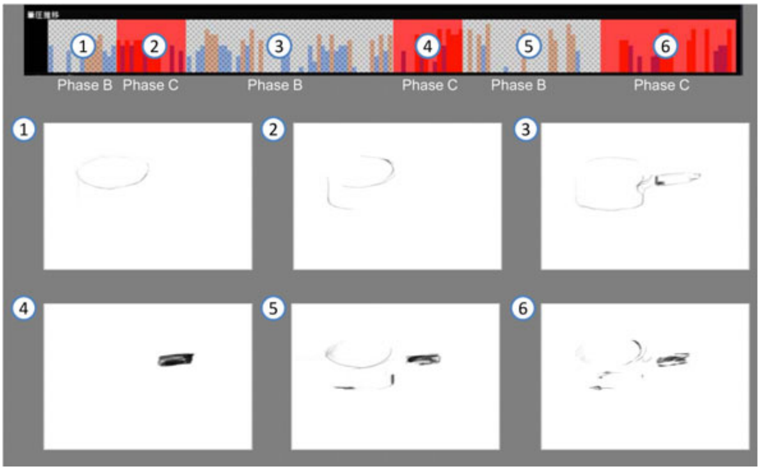


Fig. 6. The drawing results in each phase prior to the experiment [Subject-1]

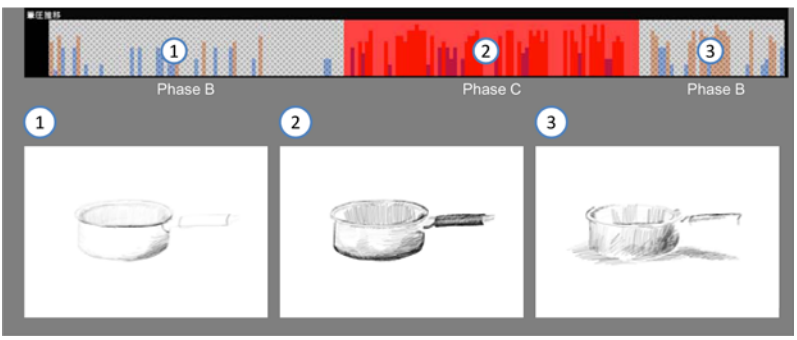


Fig. 7. The drawing results in each phase prior to the experiment [Subject-2]

After the 1 month experiment, in 6 out of 15 students whose drawing process did not include Phase-A, a section of Phase-A began to appear in their work. These students certainly used our system in their individual learning. The other 9 students could not draw an outline of the motif at the beginning of their drawing yet. The frequencies of use of our system were lower than the students who were able to improve their drawing. Fig. 8 shows the final works after the experiment. Fig. 9 and Fig.10 show the drawing results of each phase for two students (the same students as Fig. 6 and Fig. 7). The qualities of the final works are not high (almost the same level as 1 month before), but in these results, learners try to make the outline of the motif using low pressure at the beginning of their drawing. The following comments were collected from the subjects after the experiment.

- I can replay my own drawing process by using this tool, so I can carefully and repeatedly check my bad habits.
- I can view and replay the drawing processes of others, so I can better understand the necessary techniques.

– The timing of advice and the timing of bad drawing is synchronized, so I can easily recognize my drawing points which need amendment.

From these results, we can see the potential for educational effectiveness of the DPM based individual drawing learning support.

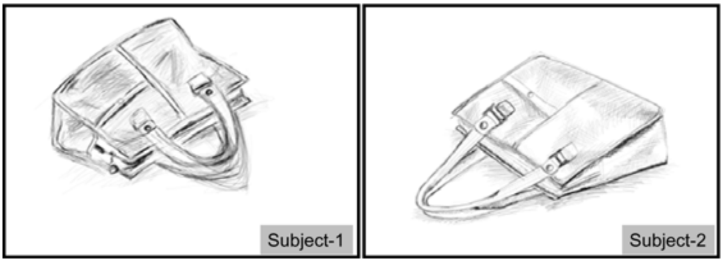


Fig. 8. The final works after the experiment

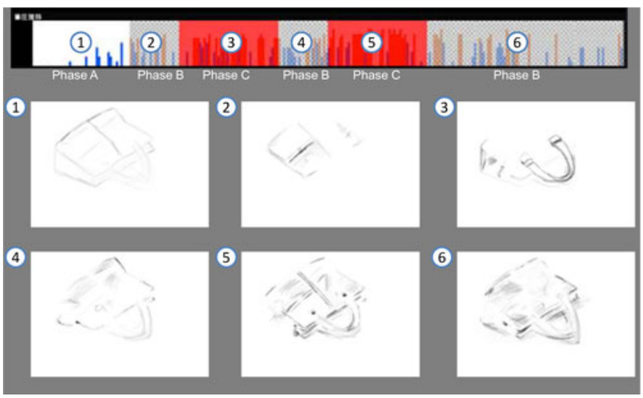


Fig. 9. The drawing results at each phase after the experiment [Subject-1]

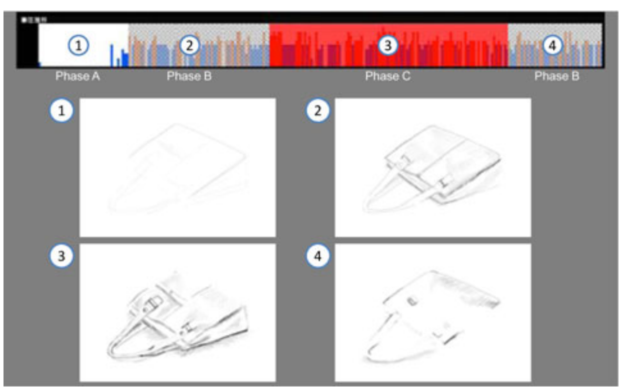


Fig. 10. The drawing results at each phase after the experiment [Subject-2]

4 Conclusion

In this paper, we describe the overview of our drawing support system in a networked environment, then the concept and the functions of the DPM are shown. Its application to classroom learning and individual learning are also considered. Finally, the educational effectiveness of our system is examined. From an experiment with students in an art school, we can see that our drawing learning support system is useful if the users are limited to beginners of drawing.

In future work, we will find adequate drawing tasks to suit constraints and limitations of the digital pen. During the operation of this system in practice, we try to arrange and revise the DPM and formalize more instructional rules for drawing learners.

References

1. Tweddle, L.K.: Reinventing Papert's Constructionism - Boosting Young Children's Writing Skills with e-Learning Designed for Dyslexics. The Electronic J. of e-Learning 6(3), 227–234 (2008)
2. Univ. of the Arts London: MA Visual Arts (Digital Arts Online), `http://www.wimbledon.arts.ac.uk/` (accessed February 28, 2011)
3. Open Dictionaly: Arts; VisualAtrs; Drawing; Education, `http://www.dmoz.org/Arts/Visual_Arts/Drawing/Education/` (accessed May 16, 2010)
4. Sekine, E.: A Trial to develop the ART SYSTEM. Art Education 6, 89–100 (1984) (in Japanese)
5. Bernstein, N.: The Co-ordination and Regulation of Movements. Pergamon Press, NY (1967)
6. Latash, M.L.: Progress in Motor Control. Bernstein's Traditions in Movement Studies, vol. 1. Human Kinetics, Urbana (1998)
7. Furukawa, K.: Skill Science. J. of JSAI 19(3), 355–364 (2004) (in Japanese)
8. Nagai, T., Kayama, M., Itoh, K.: A basic study on a drawing-learning support system in the networked environment. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 860–868. Springer, Heidelberg (2009)
9. Ernest, W.W.: The Art of Pencil Drawing, Watson-Guptill, NY (1985)

Interactions between Human and Computer Networks: EFL College Students Using Computer Learning Tools in Remedial English Classes

Ai-Ling Wang

English Department, Tamkang University,
151 Yingzhuan Road, Danshui District
New Taipei City, Taiwan
wanga@mail.tku.edu.tw

Abstract. This study aims to explore how EFL college students perceive some educational computer programs in terms of their effectiveness in learning, their user friendliness, and the role these computer programs can play in students' learning process. Online language learning was viewed from a marketing service perspective. The researcher collected data from the online remedial English classes she offered. Students taking the course are low-achieving language learners. Qualitative data were collected from online interactions between students and students and between students and the teaching assistants or the teacher, interviews with the students, questionnaires filled out and documents submitted online by the students. Data collected for the study were analyzed according to the procedure specified in the grounded theory, including data collection, note-taking, coding, memoing, sorting, and writing. Findings of this study showed that there is a gap between how low-achieving EFL college students and language teachers view computer learning programs.

Keywords: educational computer programs, online remedial English class, grounded theory.

1 Introduction

In the educational arena, research on e-learning is always done from the teacher's perspective. For example, research studies might focus on how a teacher can develop an online course and how a teacher can interact with students online [1]. This study aims to explore some issues relevant to e-learning from students' perspectives. That is, the researcher intends to investigate how students perceive online learning and how they use the computer to interact with the teacher, teaching assistants, and their fellow learners online. By doing so, the researcher expects to bridge the gap between the teacher and the student.

On the other hand, instead of using the traditional ways of hypothesis-testing in education, this study employed the grounded theory commonly used in the management field to allow significant facts to emerge and to generate possible theories. Looked at from the management perspective, education may be compared to

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 107–112, 2011.

business transaction. In this scenario, students may be viewed as customers and online instruction is the product. As the innovator, the teacher has to be aware of students' needs in order to meet their full satisfaction. That is, to bridge the gap between teacher and student, the teacher has to well capture students' needs, interests, characteristics, and their interactions online in order to "market" the "online product" [2][3].

1.1 Background of the Study

Since English proficiency has become more and more important in an age moving toward a global village, Tamkang University in Taiwan has set a threshold of English proficiency that requires its students to pass the threshold before they can graduate. For those who are not able to pass the threshold before they graduate, they are allowed to take "online tutorial English" as an alternative.

"Online tutorial English" is offered exclusively for those who are graduating and who have not passed the threshold of English proficiency test. Participants involved in this study are in different levels and different fields of study. Students may be enrolled in the class by showing their records of English proficiency test. Students taking the course meet only in the first week of the semester and the weeks before the mid-term and the final exam. For the rest of the semester, students study English online on their own. There are educational computer programs developed for online learning. The instructor, also the researcher, posted articles on the Web for students to read each week. Each article is accompanied by a PowerPoint file that list and explain some vocabulary words, idioms, and sentences. At the end of the file, there are quizzes, and students are required to do the quizzes online. Students' performance of this course is evaluated based on their mid-term and final exam scores, their responses to the online quizzes, their interactions online, and the length of time they spend online.

1.2 The Educational Computer Programs and Interactions Online

There are basically three kinds of educational computer programs used in the study, namely WebCT, Moodle, and the instructional computer platform developed by Tamkang University. All the three educational computer programs share some essential functions, namely posting, response, online discussion, submission and correction of assignments, online testing, grading, etc. At the beginning of the semester, the teacher posted the class requirements and rules, syllabi, grouping, criteria for grading, instruction to the use of educational computer program, etc. The teacher selected articles of different topics for students to read each week. The topics chosen generally covered sports, entertainment, education, culture, health, business, technologies, and some global issues. Students had to read 2 or 3 articles on a certain topic each week. Along with the articles, the teacher also posted a PowerPoint file, explaining some important vocabulary words, sentences, grammar points, and idioms and positing quizzes for students to complete online. The time students spent on the Web was recorded in the program. Students can raise their questions and post them online whenever they want and every member in the "online community" can respond to the questions. After reading the articles, students were asked to do the quizzes relevant to the articles and submit their answers online. The teaching assistants will assess their answers and assign a grade to them.

In this research setting, there are different kinds of interaction happening, namely interactions between students and students, students and teaching assistants, students and the teacher, students and the computer, and computer and computer. These online interactions can be great indicators of how students perceive the online course and how the teacher and the teaching assistants may develop an online course to meet students' needs [4].

2 Methodology

The researcher roughly followed the principles of grounded theory to conduct the research. The researcher first collected data from different sources, namely different classes of "Online English tutorial", ranging from 2008 to 2010. The qualitative data collected included online instructional materials, students' responses to the online questions, students' written interactions with the teacher, the teaching assistants, and peer students, questionnaires they filled out online, the researcher's field notes and observation, students' written statements handed in to the researcher, and the researcher's interview with students. These data were first coded in the open coding stage. In this stage, the researcher did line-by-line coding and generated some relevant codes. These descriptive codes were then conceptualized and the number of code was reduced on a conceptual coding list. In the coding process of moving from description to conceptualization, the researcher constantly compared between and among incidents, and the emerging concepts were compared with more incidents. Finally, emergent concepts were compared to each other. At the same time, the researcher also did memoing to take some theoretical notes about the conceptual connections between

Table 1. Illustration of Research Paradigm

Online remedial English course	Teacher	Chasm	Student
Instructional materials	Online reading materials and reading instruction		Topics Level of difficulty Interests Reading load
Learning tool (the computer)	WebCT Moodle School-developed computer platform		Familiarity with the program User friendly Functions available
Availability and flexibility	Online learning (asynchronous)		Time of learning Accessibility
Interactivity online	Functions to post and respond online Contact information		Channels of interacting with members of the online community
Learning outcome	Criteria for assessing students' learning outcome		Value judgment based on internal and external motivation

categories. Then, the researcher sorted the memos to allow a theoretical outline to emerge. As mentioned by Holton [5], this study aims at "explaining" students' online behaviors and their perception of online instruction, rather than just "describing" them.

Based on the coding process mentioned above, the entire research paradigm can be conceptualized and illustrated as follows:

3 Findings of the Study

Some significant points have emerged after analysis of the data. First, findings of this study showed that most of the students preferred online instruction to onsite instruction and that only a small number of older students rather than younger students would like to have onsite instruction instead. These older students claimed that they would learn more if they physically attend the class on a regular basis and listen to the teacher explaining some grammar points. Analysis through different levels of coding has shown that older students tend to value classroom learning because they might return to their graduate studies after several years' of working in a certain field. They expect to have direct contact with the teacher and to ask questions and learn from the teacher. In a sense of business transaction, they would expect to get as much as they can for what they have paid. For them, it is a privilege to be formally-educated in a school.

For younger students, they might not have work experiences before and they usually put getting a degree as their top priority. They do not quite value learning in the classroom. In a business sense, they pay in order to get a degree (product). As long as they reach the goal, the deal is done. They are required to complete their scientific or social experiment before they can get their degree. For this reason, they spend most of their time in the lab, working on their research. Online courses provide them with flexibility to manage their time, and they can choose whatever time available for them to go online to learn English. Some of them had to work in the lab on a certain day, and a fixed class schedule would cause inconvenience for them.

Second, another characteristic apparent in the participants' perception of the online course is that most of the students regard English proficiency as an important skill to master; however, they all failed to achieve a certain degree of English proficiency. Analysis of the data showed that they considered English proficiency an important skill because, in an age moving toward a global village, having a good command of English would be a plus in their forthcoming job hunting and would benefit their future career. However, several factors contributed to their failure to master English. Several students claimed that they have a heavy load in their own field of study and they do not have much time and energy to work on improving their English proficiency, especially for those who had to spend most of their time in a lab, doing scientific experiments. Others claimed that they are more inclined to scientific issues than to language and literature. No matter how they spend time learning English, they could not manage to learn the language. The way they perceive their aptitude discourages them to learn English.

Third, still another concept shared by most of the students is that although students attending the remedial class were those who did not pass the threshold of the English proficiency test, they did not learn English just because they want to pass the English proficiency test. Some of the students explained that they might need to use English in their future career, such as writing a resume or a cover letter, reading and writing academic papers, and communication in a workplace. As soon as they have a good command of English, they would have no problem passing the English proficiency test. Some of the students took preparatory English proficiency test classes before; however, they claimed that these classes were not of great help. It is because the class size is always large and the class time is always limited.

Fourth, as far as interactions online are concerned, the research data collected showed that only a small part of the students regularly posted or responded on the program system. Analysis of characteristics of these students showed that they cared about their learning of English and about the grade they will earn for this course. Most of them asked the teacher to clarify some grammatical issues they encountered while reading the articles or pointed out some typo errors found in the PowerPoint file. Others reported some technical problems, and still others would question the grade they were assigned by the teaching assistants for the assignment they turned in online. These facts might indicate that students in the remedial English class were not really motivated to learn English if there are no internal or external factors involved in their learning process.

Fifth, generally speaking, findings of the study showed that articles chosen for the course met students' learning need. Most of the students responded that they have little problem reading the articles and that the topics are of their concern. On the other hand, some of the articles with more difficult vocabulary words, such as the one with some medical terms, were also pointed out as being too difficult for low-achieving learners of English. Immediacy to the students is also part of their concern. Some of the students suggested adding "local news" to the list of topics.

Sixth, students seemed to have little problem working on the computer programs. There were only some technical issues reported by the students and these problems were generally solved soon. Students were seen satisfied with the functions available on the computer programs.

4 Conclusion and Educational Implications

Based on the theories generated from analysis of this study, the researcher captured a picture of how students perceive online remedial English courses and how they interact online. Generally speaking, students prefer to take remedial English class online because, in this case, they can have flexibility to choose the time available for them to study online. However, this study also showed that there is a gap between teacher's expectation and student's perception of online remedial English course. To bridge the gap, teachers have to take students' perception into account.

Hopefully, this study can provide language teachers who apply modern technologies to their instruction with an insight into how an online course can be best developed to meet students' needs.

References

1. Wilkinson, J.S., Wang, A.: Crossing Borders: How Cross-cultural Videoconferencing Can Satisfy Course Goals in Dissimilar Subjects. In: Tseng, P., Kwang, R., Fox, R. (eds.), pp. 109–123. World Scientific Publishing, Singapore (2007)
2. Dick, B.: Grounded theory: a thumbnail sketch (2005), http://www.scu.edu.au/schools/gcm/ar/grounded.html
3. Borgatti, S.: Introduction to grounded theory, http://www.analytictech.com/mb870/introGT.htm
4. White, C.: Language learning in distance education. Cambridge U.P, Cambridge (2003)
5. Holton, J.A.: The Coding Process and Its Challenges. In: Bryant, A., Charmaz, K. (eds.) The Sage Handbook of Grounded Theory, pp. 265–289. Sage Publications, Thousand Oaks (2007)

Proposal of Collaborative Learning Support Method in Risk Communications

Hiroshi Yajima and Naohisa Tanabe

Tokyo Denki University
2-2 Kanda-Nishiki Cho, Chiyoda-Ku, Tokyo, Japan
yajima@im.dendai.ac.jp

Abstract. In this paper, we propose the supporting method of the risk communications that use the collaborative learning. Using collaborative learning, participant of risk communication can acquire not only knowledge the participant is interested in, but also the intention and knowledge of other party who do not concern the participant's concern. In the process of collaborative learning, participants of risk communication get the mutual understanding about risks. The feature of this method is to use the WD (Write Down) form and the construction drawing of the opinion understanding made from Fishbone.

Keywords: Risk Communication, Collaborative Learning, Participant.

1 Introduction

A social risk is diversified as the information society develops, and a complex social trouble like the youth information restriction problem occurs. And the enterprise and the society are holding various risks respectively. Recently, the phenomenon in which one risk measures generates a new risk is caused. For instance, security countermeasures such as the encryption and introduction of the public key certificate for the digitalized signature cause the personal information leak such as the address and date of birth, and the risk concerning privacy is generated as a result.

Thus, requesting the combination of preferable measures ideas (optimum solution), while considering two or more risks and costs becomes very important in the situation in which correspondence to a certain risk, increases other risks.

It is finally essential to find the most suitable solution which can form an agreement among people of decision making participation. It is need in consideration of interests between people of participation to solve these problems. At the same time, not only knowledge and judgment of the expert but also opinions of participants are necessary.

Therefore, the risk communications (RC) that are the processes to do the consensus building among those with different standpoint and aspect (the stake holder and the decision-maker are included) and specialists are needed.

As the risk communications supporting tool to solve the social risk problem and the social mutual agreement problem in the information society, the multiple risk communicator (MRC) is developed[1].

It has been understood that in the process applying MRC to the large-scale, social mutual agreement problem, the participation person's prior study is important for

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 113–120, 2011.

participation person's decision making, and also the consensus building among those who take part is moreover difficult if participation person's decision making is not enough.

In this paper, to do the mutual understanding between participants smoothly, the risk study supporting method that uses cooperative study[4] is proposed. In the proposal technique, participant studies risks concerning the interested field, and this is put into writing. Participant clarifies own intention while requesting the opinion of other participants by showing the study result, and decides the optimum solution that he consents. Communications are done among participants for each participant's optimum solution, and the final mutual agreement solution is obtained. Those of the feature of the proposed method is first participant can know intentions and the finding of other participants during the risk study period, so multipronged study can be done, and secondly each participant knows other party's intention at early stages, so mutual understanding can be achieved smoothly. The utility of the proposed method is shown by the verification experiment.

2 Risk Communications Support Problem

2.1 Outline of MRC for RC

In MRC, participant support part consists of three stages of the following (a)(b)(c). [2][3]. The RC support process is shown in Fig. 1.

(a) 1stRC : Information acquisition phase for each participant to attempt clarification of self-opinion.
(b) 2ndRC : Phase in which the mutual understanding between participants is attempted.
(c) 3rdRC: Phase to plan the agreement formation between participants.

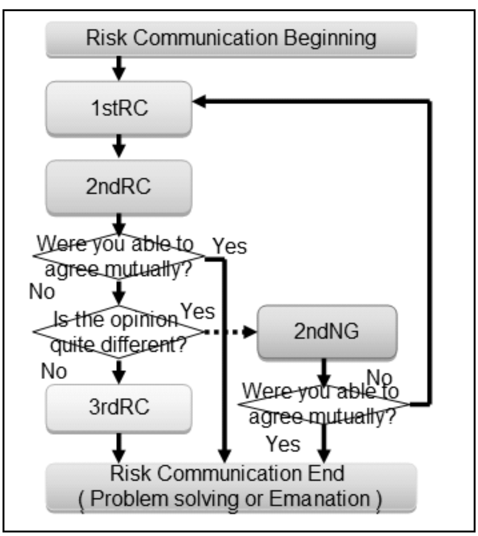


Fig. 1. RC support process

In this paper, we propose an RC support method for 1stRC and 2ndRC.

2.2 A Risk Learning Problem for the Participants

We performed an experiment. As a result of experiment, the argument between participants took much time in 2ndRC and did not advance smoothly. In 2ndRC, An argument between participants is performed so that participants find respected measures considering each other's intention. As the result, it has understood that the support of this phase is necessary for agreeing. The following issues are made clear.

Difference of the intention between participants and participant's risk understanding shortage
When participant adopts risk measures, participant emphatically take care about own risk. In that case, the participant's intention to the risk is different according to the difference of each participant's standpoint. Evaluation figure of measures was different according to this difference in each participant. (Fig.2) This phenomenon caused the discussion in 2ndRC not to go well. Because only the risk study along own intention of participant was done. Therefore these two points were made clear, that is (1)in the sturdy of participant, the important study range that is valued originally for RC was lacked. (2)Time has hung in the understanding of participant about the proof that is behind another participant's opinion.

Moreover, as for the content that became proof of the testimony of another participant, the participant's concern was left in low level.

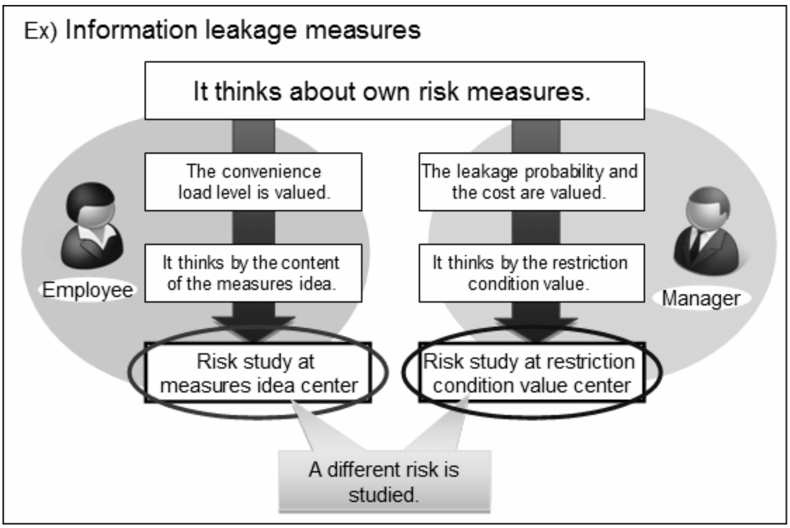


Fig. 2. Evaluation figure and risk study

Timing of information exchange and information log. In the RC support method that uses above-mentioned MRC, it is necessary to execute a lot of study processes in 1stRC. For instance, when those who take part study the information of the measures idea, participant should study "Leakage probability", "Measures cost", "Convenience load level", and "Privacy load level", etc. for the individual information leakage problem. In that case, participant study each item with the measures idea unit. Participant should study at the same time again while combining these items.

In this case, when the risk measures ideas are 15 pieces, participant should think about the combination of these measures ideas. In a word, various study cases exist for participant. Therefore, it becomes difficult for participant to integrate and understand study content that be studied in the first stage, while becoming the latter half of study.

Moreover, as the result, the chance of the information exchange in this phase was few though 2ndRC was being offered in the RC support method described in 2・1 as a place for the information exchange. This issue cause by difference of the content type that participants study and by the difference of the amount of study. In 2ndRC, participants discuss and negotiate solution based on the knowledge that they obtained in their risk study. In that case, the amount of unknown content or hearing only in the word for certain participant has increased when there is a study difference among participants. As a result, the following inconvenient cases were generated as the discussion was done repeatedly.

(1) Case where important points for participant became indefinite in the discussion
(2) Case where participant missed relativity with own risks in the discussion.

Therefore, the discussion for the consensus building was not settled well.

3 Proposed Solution

To solve the above-mentioned problem, we propose the solution that adds the viewpoint of the cooperative study to conventional RC method in this paper. This solution consists of the following three methods.

3.1 RC Using Cooperative Study

Cooperative study is based on the assumption of a close, active interaction activity between learners, and enables metacognition formation (Expression power, persuasive power, problem discovery way, problem solving way, observation method of others speech and behavior and look into oneself of self-speech and behavior) and deepen the knowledge, and gives overall view of the target to participant of RC.

One of the features is the technique called 'write down'(WD, Zaika in Japanese). In the WD(Zaika), people writes knowledge and the reproof as documents or figures, and these documents or figures are left as the log for study. Participant can review the study finding and the self-intention at any time by looking at WD. Therefore, an active discussion becomes possible by executing WD.

3.2 Cooperation Type Risk Study Method

In considering collaboration type risk study, we proposes two type of methods, that is "Allotment type study" and "Development type study".

Allotment type study. The risk is studied to the event based on own necessities of each participant. Next, the content that participants studied mutually is given each other. It aims that all participants cover about the range of study necessary for RC by this procedure.

Participant independently decides the individual participant's range of study. However, the facilitator intervenes the participant's study according to the study theme, and in this case, participants study separately mutually. In that case, the difference of the recognition between participants becomes clear by comparing the study results between participants. Moreover, the study range that all participants did not study at all is covered by what the specialist explains based on WD information. (Fig.3)

Moreover, the following effects arise so that the learner may teach the content of study to another learner in Allotment type study.

(a) Participant has a sense of responsibility for his study.
(b) Participant does an independent study action.
(c) The interpersonal relationship between Participants is formed

Allotment type study can be used as a personal cure for Epistemic Egocentrism Epistemic. Egocentrism is a bias through that the other person also knows what I know. In allotment type study, participant examines whereabouts and the bigness and smallness of an actual risk closely. Allotment type study makes participant's acknowledged risk visible, makes participant conscious of the risk, and, in addition, corrects the risk acknowledgment of participant.

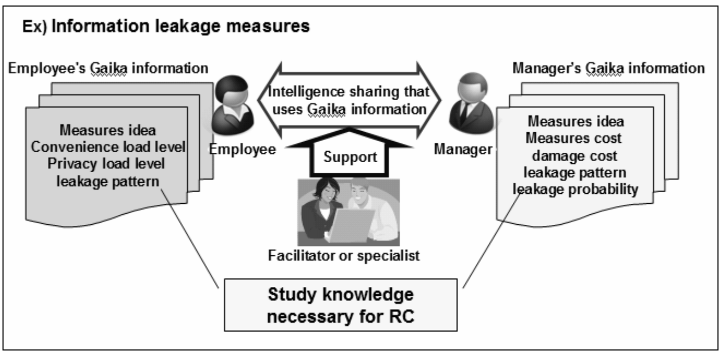


Fig. 3. Range of knowledge necessary for RC

Development Type Study. The development type study uses participant's "Opinion" and "Actual experience", etc. in WD that is the feature of the cooperative study. The chance to think about the problem, the finding, and the opinion which is not studied for myself alone, by knowing another participant's intention and finding through WD arises. In a word, deep risk understanding arises by facing the problem, the finding, and the opinion that are not studied for himself alone.

4 RC Process Using Cooperative Study

This chapter describes a concrete process procedure of the proposal method described in Chapter 3 (Fig.2).

4.1 1st Step Risk Study and Mutual Understanding

In this phase, the information gain (risk study) done with 1stRC and discussion done with 2ndRC is united in the technique. There is a feature of this phase in the point for participant to study the risk while sharing information with other participant. However, there is no place of the spoken answer discussion among participants, and mutual understanding among participants is achieved by sharing information by WD.

In this paper, the cooperation type risk study is done in the form of decentralized study., In the process, participants advance self-study (individual study) and the intelligence sharing and these studies are advanced concurrently. Two kinds of WD patterns are set. In a word, these are WD pattern by self-study, and an opinion pattern to the intelligence sharing. Moreover, the form of the development type study that participant reaches the in-depth understanding further is adopted by receiving another participant's opinion in this method.

4.2 2nd Step (Consensus Building)

This phase is placed on the extension of mutual understanding in 1st Step. In this phase, the discussion by the spoken answer among participants is chiefly performed. In this phase, the facilitator makes the construction drawing of the opinion understanding (Fig.4) based on the hope solution on each participant by putting out first of all with 1st Step and WD information. Participants discuss about final solution by using this construction drawing. The construction drawing of the opinion understanding is an improvement of Fishbone figure to understand the situation of the development type study of other participant and the intention by present quickly.

In the construction drawing for the opinion understanding, a horizontal axis is a time axis, and content of WD of participant and WD content of other participants who see the content of each participant's WD is described. In this figure, the RC name is described in the screen left end, and the first WD of the event on an upper and lower edge is described, and, in addition, final WD is described on a fat line at the center. The first WD and final WD are tied in the line, and other WDs are on the way of the line Final WD is decided from WD from other participants to the first WD, and Final WD is put out by receiving these WD. When the discussion emanates without the mutual agreement solution's to which all participant's opinions correspond or when mutual agreement solution among participant is obtained, this phase is assumed to be an end. (The above-mentioned mutual agreement solution contains the proposal of concerning alternatives about the measures idea that MRC offers and proposal and the adoption of compromise solution about the change of the measures idea setting etc.)

5 Verification Experiment

5.1 Experiment Purpose

The proposal technique is applied to the individual information leakage problem, and whether the problem described in 2・2 is solved is verified.

5.2 Precondition

In this experiment, RC intended for the security review is performed, for the enterprise that has urged by the necessity to solve the individual information leakage problem. A student in one's twenties performed the manager post and the employee as a testee. Eight students participated, and experiments on four cases are performed each by two students. The facilitator advised the testee on each phase at any time while experimenting. The following three points were required for participants.

(a) The manager and the employee take cooperated each other, standpoint for the company.
(b) Participant doesn't keep a secret to another.
(c) Participants agree on the final search of each other for the solution to satisfy.

5.3 Outcome of an Experiment

All groups reached the consensus building as a result of applying the RC support method described in Chapter 4. (Fig.4 shows the construction drawing of understanding of the opinion between participants made as a result of the experiment.)

5.4 Consideration and Finding

The proposed method to use the cooperation type study was found to be effective for smooth RC, that is, smooth selection of optimal solution and the consensus building from the outcome of an experiment. An insufficient points were observed about the risk understanding by participant and the understanding of the risk structure by participant. The information exchange at yhe timing that the risk was studied and the consensus building support by the construction drawing of the opinion understanding were effective for RC. For the WD form, the evaluation value was obtained from the testee with the high appraisal of four or more, including the following comment.

(a) Participation person's WD is easy.
(b) The understanding of other participant's WD was easy.

It was clarified that the intelligence sharing at an early stage was effective from the free description type questionnaire that had been done at the same time after experiment including the following opinions.

(1) Participant worked on RC valuing other participant's opinions.
(2) The utility of measures was able to be discussed among participants.
(3) It became easy to compromise because it was able to confirm other participant's intentions before own opinion hardened.

As a whole, the process to which the discussion for the consensus building was done from the risk study was observed, and the problem described in 2.2 was solved.

6 Conclusion

In this paper, we proposed information acquisition methods, that are the development type study that used the cooperative study, and allotment study method, as a support method of the risk communications in the cooperation type study. Mutual understanding and the consensus building supporting tools such as WD forms and Fishbone were introduced. As a result of the RC experiment, the effectiveness of the proposed method became clear.

We will develop with a more effective method by systematizing the proposal method in the future.

References

[1] Sasaki, R., Hidaka, Y., Moriya, T., Taniyama, M., Yajima, H., Yaegasi, K., Yoshiura, Y.: Development and application of the multiple risk communicator's. Journal of Information Processing of Japan 49(9), 3180–3190 (2008)
[2] Watanabe, T., Yamamoto, H., Yajima, H., Sasaki, R.: Evaluation of participant information acquisition support method in multiple risk communicator. IEEJ Trans. EIS 128(2), 310–317 (2008)
[3] Matsumoto, S., Hiroshi, Y., Sasaki, R.: Proposal of smooth risk communications support method in multiple risk communicator. In: Symposium on Cryptography and Information Security, SCIS (2006)
[4] Miyake, N.: Cooperated study and AI. Journal of the Japan Society of Artificial Intelligence 23(2), 174–183 (2008)

Evaluation of Online Handwritten Characters for Penmanship Learning Support System

Tatsuya Yamaguchi[1], Noriaki Muranaka[2], and Masataka Tokumaru[2]

[1] Graduate School of Kansai University
3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
[2] Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
{k526329,muranaka,toku}@kansai-u.ac.jp

Abstract. This paper proposes a method for the evaluation of online handwritten characters for a penmanship learning support system. In Japan, many people desire to write beautiful characters, as evidenced by the fact that correspondence courses on penmanship are very popular. However, correspondence courses lack real-time feedback because of the time required for sending materials, namely, the learner's copy and the teacher's correct result. Therefore, we have developed a penmanship learning support system that automatically evaluates the characters that a learner writes using a personal digital assistant (PDA). We examined the validity of our proposed character-evaluation method, and proved it to be valid.

Keywords: Penmanship, Character Evaluation, PDA, Similarity, Balance, Speed, Advice.

1 Introduction

In recent years, many people have begun to use the personal computer (PC) in their daily lives, because its performance has improved and its price has declined. Therefore, various learning support systems that use the PC have been studied [1,2]. Among these studies, support systems that promote the learning of calligraphy and penmanship have been actively studied [3,4]. One reason for such interest is that there are many people who desire to write beautiful characters. In Japan, it is said that handwritten characters indicate one's personality. However, many people do not write characters frequently because of the diffusion of PCs and mobile phones, although the Japanese still write important documents, letters, and resumes by hand to communicate their feelings more precisely. Therefore, many people are learning penmanship. The methods for learning penmanship include correspondence courses and classes. In correspondence courses, learners learn the shape and balance of a character on the basis of a model character in the text and its accompanying explanation, at their convenience. Then, the learner sends his/her copy of the character to a teacher, and the teacher corrects it and sends the result back to the learner. Thus, correspondence courses lack real-time feedback because of the time

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 121–130, 2011.

required for sending materials. Attending a classroom course is the best way to learn penmanship because of the direct interaction between teachers and learners. However, because learners must attend penmanship classes regularly, classroom courses are not suitable for busy people. Therefore, we have developed a penmanship learning support system that automatically evaluates the characters a learner writes using a personal digital assistant (PDA) or a liquid-crystal pen display. The purpose of this system is to reproduce the environment of a penmanship classroom at home. Furthermore, learners can learn penmanship using a palm-sized PDA without being restricted for place or time. This system evaluates the characters a learner writes by comparing the features of model characters with those of the characters written by the learner. By evaluating these features, learners begin considering the features of the characters. We had proposed a feature-extraction method in a previous work [5]. By extracting the features on the system, we can apply one character-evaluation method to all characters. In other words, it is not necessary to prepare a separate character-evaluation method for each character. This paper describes the features of this character-evaluation method.

2 System Overview

This system uses two types of characters: model characters and learner’s characters. Model characters are the characters a penmanship teacher has written beforehand. Learner’s characters are those a learner writes when he/she learns. Learners learn the shape and balance of model characters by copying them. The learning processes in the system are similar to classroom techniques of penmanship. First, learners practice by tracing a model character. Next, they make a fair copy of it. Then, the system evaluates their character and presents advice. The learners review the advice and repeat this process.

2.1 Hardware

The authors implemented this system on a PDA (Fig. 1). The PDA features a palm-size coordinate input device that provides easy synchronization between a PC and the PDA.

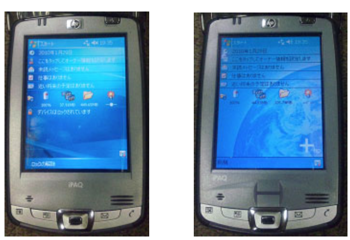
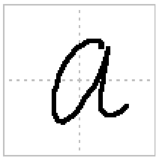


Fig. 1. Left: HP iPAQ hx2400, Right: HP iPAQ hx2700

2.2 Character Data

The character data has coordinates at intervals of 10 ms, an example of which is shown in Fig. 2(b). These intervals enable the reproduction of the writing process (the movement of the pen point) of the character. The data text consists of the index, take-off and landing of the pen point (take-off: 0, landing: 1), stroke count, abscissa, and ordinate (from the left). The index is the time elapsed between reaching the coordinates and starting to write a character.



0	1	1	114	78
1	1	1	115	78
4	1	1	115	77
6	1	1	115	76
7	1	1	114	75
9	1	1	113	74

(a) (b)

Fig. 2. For a character "a" written as (a), the system saves a data text of the character; (b) shows a part of the data text

2.3 Function

This system has four types of functions, namely, animation, trace, evaluation, and advice.

The animation function reproduces the writing process of a character that a teacher and learner write. This is the biggest advantage of device learning because paper learning tends to lose this process. Furthermore, this function enables the determination of the speed at which a character is written.

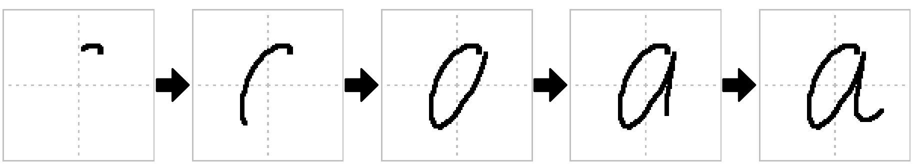


Fig. 3. Reproduction image of the animation function

The trace function of this system enables the tracing of the model character. In correspondence courses and classes on penmanship, a learner practices initially by tracing model characters.

The evaluation function evaluates the learner's character on a scale of 0 to 100. The elements of evaluation include similarity, balance, and speed of the character. This paper describes the details of the evaluation method in Section 3.

The presentation of advice is divided into two categories: an accompanying explanation and a correction. Accompanying explanation describes how to write a beautiful character before you write it. Correction teaches what needs to be corrected in the character after you write it. This system uses this advice, and consequently learners progress their penmanship learning. In this paper, the advice function indicates a correction.

2.4 GUI

The GUI in this system contains a learning window, a result window, and an advice window for each evaluation element.

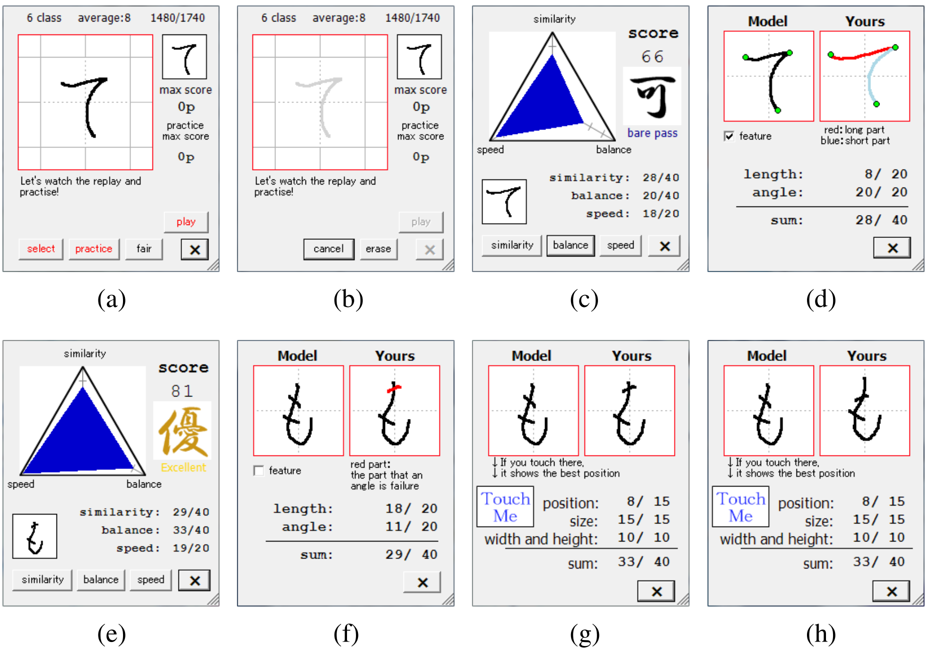


Fig. 4. Windows of the penmanship learning support system. (a) Learning window. The character in the upper-right corner of the window is a model character. In this window, learners can reproduce the writing process, practice, and make a fair copy of a model character. (b) Practice window. In this window, learners can trace the model character. (c) and (e) Result windows. (d) and (f) Advice windows regarding similarity. (g) and (h) Advice windows regarding balance. (g) is the normal window. When a learner touches “Touch Me” on (g), the window becomes (h), which shows the best position of each stroke.

3 Character Evaluation

This system evaluates a handwritten character on a scale of 0 to 100 on the basis of the evaluation elements of similarity, balance, and speed. Fig. 5 shows the itemized evaluation elements and their maximum scores.

$$
total \quad (100p) \left\{ \begin{array}{ll}
similarity \quad (40p) & \left\{ \begin{array}{ll} length & (20p) \\ angle & (20p) \end{array} \right. \\[2ex]
balance \quad (40p) & \left\{ \begin{array}{ll} position & (15p) \\ size & (15p) \\ width\,and\,height & (10p) \end{array} \right. \\[2ex]
speed \quad (20p) & \left\{ \begin{array}{ll} speed\ rate & (10p) \\ speed & (10p) \end{array} \right.
\end{array} \right.
$$

Fig. 5. Allotment of points. Values in parentheses are the maximum scores. The total score is 100 points

3.1 Similarity

Similarity is evaluated on the basis of the lengths and angles between the features of a character. The feature-extraction method is described in [5]. After the system extracts the features of each character, it maps the features of the model character and learner's character (Fig. 6). The mapping method is described in [6]. The length and angle are evaluated from the lengths and angles between the mapped features in a row using Table 1 and the following equations:

$$Lr_i = \sqrt{30/Lm_i}\,, \tag{1}$$

$$Ar_i = \sqrt{Lm_i \times Ll_i \Big/ \sum_{i=1}^{n} Lm_i}\,, \tag{2}$$

$$L_{ave} = \frac{1}{n}\log_2\left[\sum_{i=1}^{n} Ll_i \Big/ \sum_{i=1}^{n} Lm_i\right], \tag{3}$$

$$Ld_i = \left[Lr_i \times \left(Lm_i - Ll_i / L_{ave}\right)\right]^2, \tag{4}$$

$$Ad_i = \left[Ar_i \times \left(Am_i - Al_i\right)\right]^2, \tag{5}$$

$$Ld = \frac{1}{n}\sum_{i=1}^{n}\begin{cases}(Ld_i + 20)/5 & (Ld_i > 5) \\ Ld_i & (otherwise)\end{cases}, \tag{6}$$

$$Ad = \frac{1}{n}\sum_{i=1}^{n}\begin{cases}(Ad_i + 4)/5 & (Ad_i > 1) \\ Ad_i & (otherwise)\end{cases}. \tag{7}$$

In Table 1 and equations (1)–(7), L: length, A: angle, m: model character, l: learner's character, r: ratio of the model character and the learner's character, d:

difference, i: index, and n: total section number. The scores of the length and angle are given by the points deducted, depending on the solution of (6) and (7), from 20.

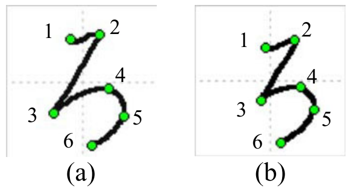


Fig. 6. Mapping points of the features in each character. (a) Model character. (b) Learner's character. Circles denote the features of each character.

Table 1. Length and angle of the features mapped in Fig. 6. Section is the area between the character's features, Model is the length and angle in a section of the model character, and Learner's is the length and angle of the learner's character.

(a) Length

Index (i)	Section	Model (Lm_i)	Learner's (Ll_i)
1	1--2	19.2	23.2
2	2--3	59.2	53.1
3	3--4	39.4	31.3
4	4--5	20.6	13.9
5	5--6	28.3	37.4

(Unit: pixel)

(b) Angle

Index (i)	Section	Model (Am_i)	Learner's (Al_i)
1	1--2	9.0	10.3
2	2--3	239.5	235.4
3	3--4	24.0	19.7
4	4--5	299.1	324.0
5	5--6	222.1	235.9

(Unit: degree)

3.2 Balance

The system evaluates the proportion of each stroke. Balance is evaluated on the basis of position, size, and width and height.

Position. The position of each stroke is the average of the x and y coordinates in the coordinate data that composes the stroke. It calculates the variance among the subtractions of the model's x coordinates from those of the learner in each stroke. Similarly, the variance in the y coordinates is calculated. The score of position is given by the points deducted, depending on the square root of the sum of the variances in the x and y coordinates, from 15.

Size. The stroke size is decided by the width and height in the stroke, as follows:

$$stroke\ size = \begin{cases} width + height/width & (width > height) \\ height + width/height & (otherwise) \end{cases}. \tag{8}$$

The score of the size is given by the points deducted, depending on the variance of the stroke sizes, from 15.

Width and Height. σ_W and σ_H are calculated from the following equations:

$$ave = \frac{1}{m}\sum_{i=1}^{m}\log_2(L_i / M_i), \tag{9}$$

$$\sigma = \frac{1}{m}\sum_{i=1}^{m}\{[\log_2(L_i / M_i) - ave] \times |L_i - M_i| \times M_i / 50\}^2 . \tag{10}$$

In equations (9) and (10), L_i: width and height in the *i*th stroke of a learner's character, M_i: width and height of the model character; and *m*: total stroke number of the character. If $m = 1$, σ is calculated using the following equation:

$$\sigma = |L_i - M_i| \times M_i / 50 . \tag{11}$$

The scores of the width and height are given by the points deducted, depending on P, from 10. P is calculated using the following equation:

$$P = \sqrt{\sigma_W{}^2 + \sigma_H{}^2} . \tag{12}$$

3.3 Speed

The system evaluates whether a character is written with a well-modulated speed. Speed is evaluated on the basis of the speed rate and the time rate.

Speed Rate. The speed rate is evaluated by the following evaluation method. First, it divides the model character into several parts at time intervals of 200 ms. Next, it divides the learner's characters into part numbers of the model character. Then, it calculates the speed in each section of the characters and the relative ratio of each section speed of the model character and the corresponding section speed of the learner's character. The score of the speed rate is the number proportional to the sample variance of the relative ratio, subtracted from 15. The following is a method for characters that consist of one stroke and a non-turn in the stroke. If these characteristics exist, the system divides the characters at those points. Then, it calculates the sample variance of the relative ratio in each section, and the weighted average. The weight is the sample size. In this case, the score of the speed rate is given by the points deducted, depending on the weighted average, from 10.

Time Rate. The time rate evaluates the relationship mentioned in the above section. The system calculates the time it takes to write each section of the model character and the learner's character, and the relative ratio of these characters in each section. The score of the time rate is given by the points deducted, depending on a sample variance of the relative ratio, from 10.

4 Experiment and Results

We examined whether the use of these evaluation elements in the system improved the efficiency of penmanship learning by testing it with ten university students. The examination spanned two weeks, for approximately 10 min per day. We provided PDAs that implement this system to the students, and used 46 Japanese "Hiragana" characters as model characters. We included a "Report button" in the system, so that a student could report a disagreement with the evaluation result by pushing the button. In reporting, a student selects a disagreed element among three evaluation elements (possible multiple-choice answers or non-choice answer). The evaluation result information is then saved in the PDA. After the examination period, we collected the PDAs and distributed a questionnaire (Table 2) to the students. Table 3 shows the experimental results.

Table 2. The questionnaire distributed to the students, with a scale of 1 to 5: 1. Strongly disagree, 2. disagree, 3. neither agree nor disagree, 4. agree, 5. strongly agree.

Q. 1	The evaluation method is valid.
Q. 2	The model characters are beautiful.
Q. 3	The more you use the system, the more you can write beautiful characters.

Table 3. Selected frequency for disagreement of each evaluation element (Similarity: S, Balance: B, Speed: S, and Non-choice: N), total report frequency (TRF), total learning frequency (TLF), and questionnaire (Table 2) results of the students

Student	S	B	S	N	TRF	TLF	Q. 1	Q. 2	Q. 3
A	3	2	12	3	16	1112	2	1	3
B	1	0	1	0	2	198	4	5	4
C	0	0	0	0	0	180	5	4	4
D	0	0	1	0	1	392	4	4	4
E	0	0	4	0	4	701	4	3	4
F	0	0	1	0	1	960	2	3	4
G	0	0	0	0	0	2226	5	5	4
H	0	0	0	0	0	368	2	2	3
I	1	1	1	2	4	179	4	5	4
J	9	16	17	0	22	151	2	2	3
Average	1.4	1.9	3.7	0.5	5	647	3.4	3.4	3.7

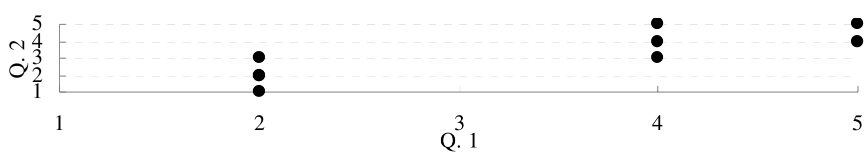


Fig. 7. Scatter diagram showing the questionnaire results of each student for Q. 1 and Q. 2

Fig. 8. Scatter diagrams showing the averaged evaluation results of each character in terms of the number of strokes, the averaged maximum score of each evaluation element, and total

5 Discussion

The evaluation method was proved valid because the total report frequency was 50, and seven out of ten students evaluated Q. 1 concerning validity as 4 or more (Table 3).

Speed was the most reported among the three evaluation elements because advice presentation regarding speed evaluation was provided to the students. In the evaluation of similarity and balance, the system indicates the incorrect part in color. In terms of speed, it reproduces model characters and learner's characters at the same time. The system needs to provide better visual advice.

Fig. 7 shows the correlation between the validity of the evaluation method and the beauty of model characters. It indicates that validity is decided by whether a student likes the model character. There are many types of beautiful characters; therefore, the model character of the system should be a character the learner likes.

Fig. 8 shows the relationship between the number of strokes in each character and the score of each evaluation element. The scores of the characters of one of the strokes are large, because these characters are the simplest. However, others show no correlation. This proves that the evaluation method is not dependent on the number of strokes in each character.

6 Conclusion and Future Work

We proposed a character-evaluation method and examined its validity. The method was proved valid because there were only a few student disagreements and seven out of ten students evaluated the questionnaire concerning validity as 4 or more. To obtain the agreement of more students, we need to provide visual advice of speed evaluation and prepare various model characters that the students like. In our future work, we propose providing visual advice and extending the system's experimentation with more people.

References

1. Miura, M., Kunifuji, S.: Hybrid approach of augmented classroom environment with digital pens and personal handhelds. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) KES 2006. LNCS (LNAI), vol. 4252, pp. 1019–1026. Springer, Heidelberg (2006)
2. Park, K.S., Leigh, J., Johnson, A.E., Carter, B., Brody, J., Sosnoski, J.: Distance learning classroom using Virtual Harlem. In: 7th International Conference on Virtual Systems and Multimedia, Berkeley, pp. 489–498 (2001)
3. Ichimura, H., Suzuki, M., Murai, M., Yamashita, S., Kuroda, M., Kuroiwa, T., Yoshida, K., Yazawa, S., Mizuno, T., Sakai, S.: Design of Next Generation Distance-Learning System for Penmanship and Calligraphy. ACIS International Journal of Computer and Information Science 2(4), 162–171 (2001)
4. Yamasaki, T., Hattori, T.: Computer Calligraphy: Brush Written Kanji Formation Based on the Calligraphic Skill Knowledge. IEICE Transactions on Information and Systems E80-D 2, 17–175 (1997)
5. Yamaguchi, T., Muranaka, N., Tokumaru, M.: Penmanship Learning Support System: Feature Extraction for Online Handwritten Characters. In: 9th International Conference on Entertainment Computing, Entertainment Computing - ICEC2010, Seoul, pp. 496–498 (2010)
6. Sawamoto, T., Muranaka, N., Tokumaru, M., Imanishi, S.: Construction of the penmanship learning support system—Examination about the features extraction and mapping during model character and handwriting character. IEICE Tech. Rep., 109(335), 89–94 (2009) (in Japanese)

Facial Expression Recognition for Learning Status Analysis

Mau-Tsuen Yang, Yi-Ju Cheng, and Ya-Chun Shih

National Dong-Hwa University, Hualien, 974, Taiwan
mtyang@mail.ndhu.edu.tw

Abstract. Facial expression provides an important clue for teachers to know the learning status of students. Thus, vision-based expression analysis is valuable not only in Human-Computer Interface but also in e-Learning. We propose a computer vision system to automatically analyze learners' video to recognize nonverbal facial expressions to discover learning status of students in distance education. In the first stage, Adaboost classifiers are applied to extract candidates of facial parts. Then spatial relationships are utilized to determine the best combination of facial features to form a *feature vector.* In the second stage, each feature vector sequence is trained and recognized as a specific emotional expression using Hidden Markov Model (HMM). The estimated probabilities of six expressions are combined into an *expression vector*. The last stage is to analyze the expression vector sequence to figure out the learning situation of the student. Gaussian Mixture Model (GMM) is applied to evaluate three learning scores (Understanding, Interaction, and Consciousness) that are integrated into a *status vector*. Each evaluated status vector reflects the learning status of a student and is helpful to not only teachers but also students for improving teaching and learning.

Keywords: Facial expression recognition, Learning status analysis.

1 Introduction

Most existing e-Learning systems focus on the use of instructor's video. However, learners' videos are critical for instructor to know the learning status of students. Advances in computer processing speed and imaging technology make it possible to capture each learner's video and estimate learning status using low-cost hardware (off-the-shelf PC and webcam). The estimated learning statuses provide valuable information for bi-directional online interaction and distance learning evaluation.

In order to automatically discover learning status from learner's video in real time, we propose a facial expression recognition system for learning status analysis. Fig. 1 shows the flowchart of the proposed system. In the first stage, Adaboost classifiers are applied to extract candidates of facial parts. Then spatial relationships are utilized to determine the best combination of facial features to form a five-dimensional vector called *feature vector.* In the second stage, each feature vector sequence is trained and recognized as a specific emotional expression using Hidden Markov Model (HMM). The estimated probabilities of six expressions (Blink, Wrinkle, Shake, Nod, Yawn,

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 131–138, 2011.

and Talk) are combined into a six-dimensional vector called *expression vector*. The last stage is to analyze the expression vector sequence to figure out the learning situation of the student. Gaussian Mixture Model (GMM) is applied to evaluate three learning scores including understanding, interaction, and consciousness. The estimated scores are integrated into a three-dimensional vector called *status vector*. The status vector reflects the learning status of a student and is helpful to not only teachers but also students for improving teaching and learning.

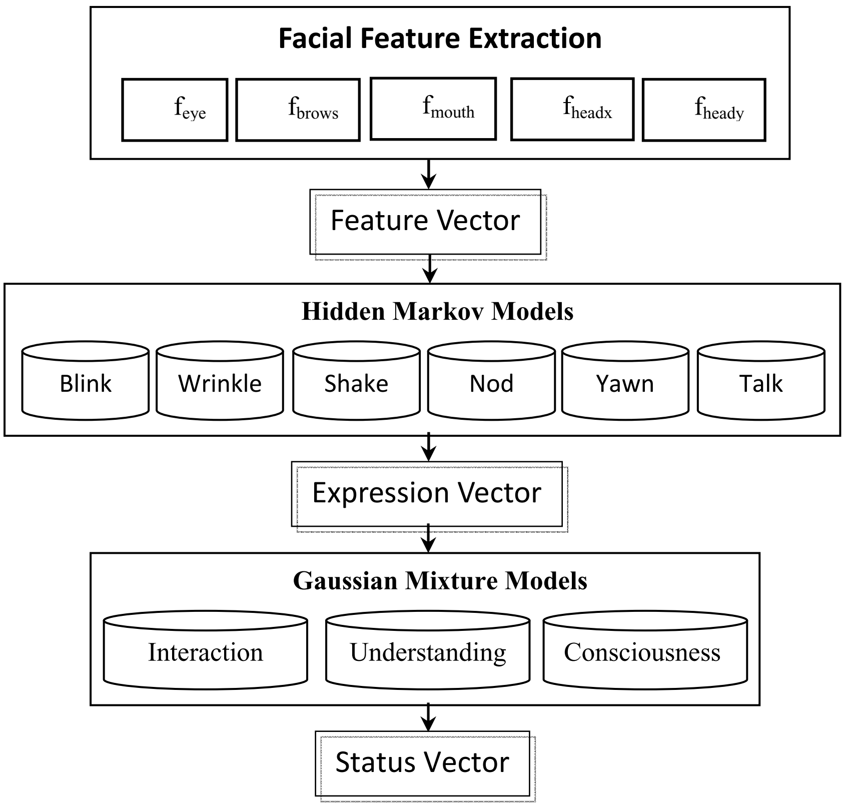


Fig. 1. Flowchart of the proposed system

2 Background

The research of face detection carries on for decades and varies from intuitional color-based, shape-based to multimodal feature-based approaches. Kumar and Bindu [8] utilized YCbCr color space to find face. Seo et al. [15] combined Snake with color space to improve accuracy of face detection. Saatci and Town [14] used Active Appearance Model (AAM) for face detection but it is only effective for specific trained person. Rowley et al. [13] applied neural network to detect face but an individual neural network need to be trained for each head orientation. Viola and

Jones [17] presented a face detection system called AdaBoost that constructs a strong classifier by linking a boosted cascade of simple haar classifiers. AdaBoost achieved rapid and stable face detection so that it is widely used for face detection nowadays. Moreover, Peng et al. [23] combined color information with AdaBoost in order to increase the robustness. Cristinacce and Cootes [22] integrated AdaBoost and shape constrains to improve the performance of face detection. Wilson and Fernandez [20] refined the AdaBoost algorithm to address the problem of head rotation.

Facial expression recognition can be performed on either a static image or an image sequence. Several methods only rely on spatial information in a still image to recognize facial expressions. Saatci and Town [14] used Support Vector Machine (SVM) to separate difference expressions by estimating hyper planes. Besides, gender classifier is applied to improve the expression recognition performance. Liu et al. proposed Fusion Neural Network (FNN) [10] to integrate multiple Gabor features for expression recognition. They also utilized Gaussian Mixture Models (GMM) to analyze eigenvector of facial expressions [9]. Cheon and Kim [4] integrated K-Nearest Neighbors (KNN) and manifold learning to classify different facial expressions. Cao and Tong [3] proposed spatial Embedded Hidden Markov Model (EHMM) to recognize facial expressions. Jung et al. [7] used AdaBoost algorithm to classify expression and showed the practicability of AdaBoost for not only detection but also recognition.

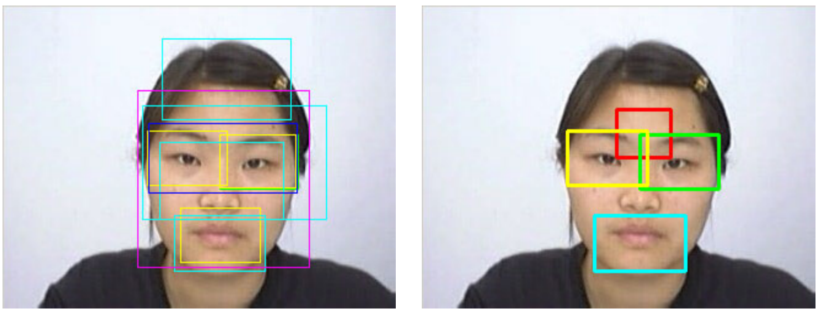
In addition to the spatial information in an image, temporal information extracted from an image sequence also provides valuable clues for expression recognition. Datcu and Rothkrantz [5] compared and showed that the expression recognition performance based on video sequences is much better than those based on still images. Ofli et al. [12] constructed several temporal parallel HMM to recognize different expressions. Wang and Ju [19] proposed a hybrid model which combined KNN with HMM to increase the recognition accuracy.

Recent advances in computer vision technology gradually reveal its potentials on e-Learning improvement. Walczak et al. [18] proposed a VR-based framework of network service for distance education. It constructed a virtual classroom in that students can learn by interacting with the virtual objects in the classroom. Calvi et al. [21] developed an eye tracking device to find learner's interest and make a warning if the learner missed his attention. Loh et al. [11] proposed an expression recognition system for e-Learning based on Gabor wavelet and Neural Network. However, their database contains only static images with four facial expression (neutral, sleepy, confuse, and smile). Relatively, we propose a facial expression recognition system working on real-time video. Based on the recognized facial expression, we further evaluate three learning scores including understanding, interaction, and consciousness to encourage online interaction and enhance the quality of e-Learning.

3 Feature Extraction

In the stage of facial feature extraction, five AdaBoost classifiers [17] are applied to detect the candidates of facial parts including face, mouth, left eye, right eye, and the region between eye brows (as shown in Fig. 2(a)). After locating these feature candidates, a relational probability function which models the spatial relationships among facial parts is defined to pick and choose the most possible combination of facial feature candidates that can reasonably form a face (as shown in Fig. 2(b)). Moreover, the optical flows of these features are calculated to estimate the head motion in both horizontal and vertical directions. The extracted facial features are integrated into a sequence of five-dimensional *feature vectors*:

$$Fv(t) = \begin{bmatrix} f_{eye}(t) \\ f_{brows}(t) \\ f_{mouth}(t) \\ f_{headx}(t) \\ f_{heady}(t) \end{bmatrix} \tag{1}$$



(a) (b)

Fig. 2. Facial feature extraction. (a) candidates of facial parts detected by AdaBoost. (b) the best combination of feature candidates considering spatial relationship among facial parts.

4 Expression Recognition

In the expression recognition stage, Hidden Markov Model [1] is adopted to recognize facial emotional expressions. The extracted feature sequences are used as the input data to train the parameters of HMM for each expression using Baum-Welch algorithm. Then Forward algorithm is used to evaluate the maximal probability of each expression. The probabilities of six expressions (Blink, Wrinkle, Shake, Nod, Yawn, and Talk) are combined into a sequence of six-dimensional *expression vectors*:

$$Ev(t) = \begin{bmatrix} p_{Blink}(t) \\ p_{Wrinkle}(t) \\ p_{Shake}(t) \\ p_{Nod}(t) \\ p_{Yawn}(t) \\ p_{Talk}(t) \end{bmatrix} \tag{2}$$

5 Status Estimation

In the stage of learning status analysis, Gaussian Mixture Model [2] is used to estimate the learning status by analyzing the expression vector sequence. Since the learning status is a lasting state, it is discovered by observing the sequence of the expression vectors for a period of time. In a fixed time interval (30 seconds in our experiments), the number of times that the expression probability is higher than a threshold is counted and record in a counting vector. EM algorithm [6] is applied to train the GMM parameters for each kind of learning status based on the counting vectors. Then the trained GMM is used to estimate the probability of respective learning status. Three learning scores (Understanding, Interaction, and Consciousness) are outputted as a sequence of three-dimensional *status vectors*:

$$Sv(t) = \begin{bmatrix} p_{Interaction}(t) \\ p_{Understanding}(t) \\ p_{Consciousness}(t) \end{bmatrix} \tag{3}$$

6 Experiments and Applications

Our preliminary experiments show that AdaBoost facial feature extraction combined with spatial relationship filtering can locate the facial features robustly. Also, the trained facial expression HMMs can recognize facial expressions reliably even for untrained people. It should be noted that it is possible to recognize several expressions with the same sequence because several expressions can appear concurrently. For example: one could blink and talk at the same moment. Finally, the GMMs can effectively estimate three learning scores: interaction, understanding, and consciousness.

It is very important for a teacher to know the learning status of students in order to improve the learning effectiveness. The proposed system has been implemented to enhance an online Virtual English Classroom called VEC3D [16]. Instructor as well as each distant student uses his or her own computer and webcam to login VEC3D online as a virtual avatar. A live facial image can be transmitted and shown above each avatar in the virtual classroom. Three bars above each facial image indicate the estimated learning scores (Interaction, Understanding, and Consciousness) of the student. In case of network with limited bandwidth, the transmission of facial images can be disabled but the learning status bars can still be shown.

The proposed system can help teachers to realize learning status of students in different teaching activities such as lecturing, group discussion, and role playing. The system estimates and records the learning status of students in class so that the teacher can adjust the teaching materials accordingly. If the *understanding* scores are too low, then the teaching materials are too hard for the students and the difficulty of the teaching material can be reduced or some remedial classes can be arranged. Besides, the teacher can record the *consciousness* scores in class and make a long term observation of learning attitude of individual student. Moreover, the *interaction* scores can evaluate the degree of interactive teaching when a teacher is giving a lecture in class. These estimated scores not only improve teaching and learning but also help fellowship development among classmates. Classmates and teachers can give their concerns to those who have low understanding, consciousness, or interaction scores.

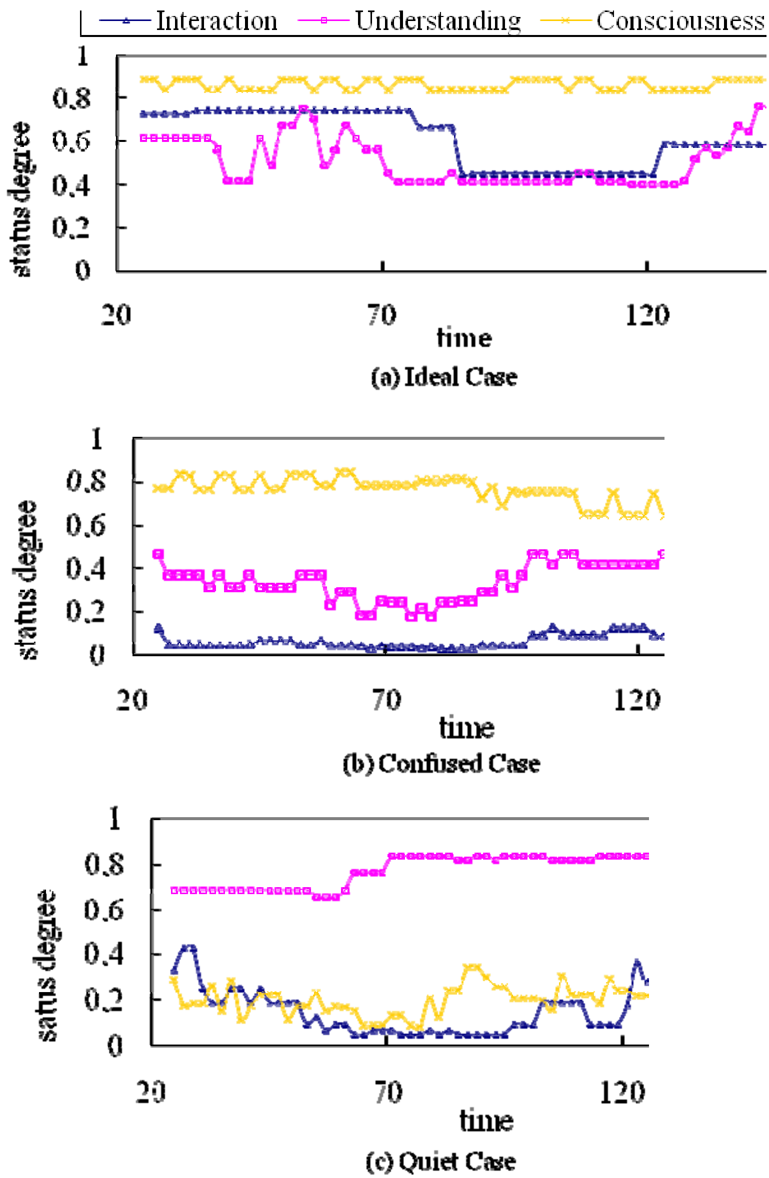


Fig. 3. Three typical learning situations when a group discussion is taking place. (a) Ideal case (b) Confused case (c) Quiet case.

Fig. 3 demonstrates three typical cases when a group discussion is taking place. The three plotted lines (blue, red, and yellow) describes the three average learning scores (Interaction, Understanding, and Consciousness) of several students in the same group. Fig. 3(a) shows ideal learning scores in a group discussion, Fig. 3(b) represents that most students in the group are confused (low Understanding score). Fig. 3(c) indicates that most students in the group are too quiet (low Interaction score). The latter two cases lead to low learning effectiveness and the teacher can get involved in their discussion and guide the students accordingly.

7 Conclusions

We propose a vision-based facial expression recognition system to improve e-Learning. We first adapt AdaBoost to extract five facial features: eye, mouth, brows, and head horizontal/vertical motion. Next, we utilize HMM to recognize the six facial expressions: blink, wrinkle, shake, nod, yawn, and talk. Finally, we apply GMM to evaluate three learning status scores: interaction, understanding, and consciousness. These evaluated scores reflect learning status and provide reference data for those who join the teaching activity directly or indirectly. According, both the instructor and the students can make improvements to increase learning effectiveness.

Our experiments show that combining Gaussian probability function with AdaBoost facial feature extraction robustly locates five facial features. Moreover, the trained HMMs reliably recognize six facial expressions among different people. Furthermore, the GMMs effectively estimate three learning scores: interaction, understanding, and consciousness. The three estimated scores can help teachers to realize learning status of students in different teaching activities. The proposed system profits to not only teachers and students for improvement purpose, but also students' parents and academic departments for evaluation purpose. Besides, it is flexible to add more facial features and expressions to make the system more accurate and trustworthy.

Acknowledgements. This work was supported by the National Science Council, Taiwan.

References

1. Baum, L., Petrie, T., Soules, G., Weiss, N.: A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains. Annals of Mathematical Statistics 41, 164–171 (1970)
2. Bilmes, J.: A Gentle Tutorial of the EM algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Model. ICSI Technical Report TR-97-021 (May 1997)
3. Cao, J., Tong, C.: Facial expression recognition based on LBP-EHMM. IEEE Congress on Image and Signal Processing 2(27-30), 371–375 (2008)
4. Cheon, Y., Kim, D.: A Natural Facial Expression Recognition Using Differential-AAM and *k*-NNS. In: IEEE International Symposium on Multimedia, pp. 220–227 (December 2008)
5. Datcu, D., Rothkrantz, L.: Facial Expression Recognition in still pictures and videos using Active Appearance Models. A comparison approach. In: ACM International Conference on Computer System and Technologies (2007)

6. Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B 39(1), 1–38 (1977)
7. Jung, S., Kim, D., An, K., Chung, M.: Efficient Rectangle Feature Extraction for Real-time Facial Expression Recognition based on AdaBoost. In: IEEE International Conference on Intelligent Robots and Systems, pp. 1941–1946 (August 2005)
8. Kumar, C., Bindu, A.: An Efficient Skin Illumination Compensation Model for Efficient Face Detection. In: IEEE Annual conference on Industrial Electronics, pp. 3444–3449 (November 2006)
9. Liu, W., Lu, J., Wang, Z., Song, H.: An Expression Space Model for Facial Expression Analysis. IEEE Congress on Image and Signal Processing 2, 680–684 (2008)
10. Liu, W., Wang, Z.: Facial Expression Recognition Based on Fusion of Multiple Gabor Features. In: IEEE International Conference on Pattern Recognition (2006)
11. Loh, M., Wong, Y., Wong, C.: Facial Expression Recognition for E-learning System using Gabor Wavelet & Neural Network. In: International Conference on Advanced Learning Technologies (2006)
12. Ofli, F., Erzin, E., Yemez, Y., Tekalp, A.: Estimation and Analysis of Facial Animation Parameter Patterns. In: IEEE International Conference on Image Processing 2007, vol. a, pp. IV293–IV296 (October 2007)
13. Rowley, H., Balujia, S., Kanade, T.: Neural Network-Based Face Detection. IEEE Journal of Pattern Analysis and Machine Intelligence 20(1), 23–28 (1998)
14. Saatci, Y., Town, C.: Cascaded Classification of Gender and Facial Expression using Active Appearance Models. In: IEEE International Conference in Automatic Face and Gesture Recognition (2006)
15. Seo, K., Kim, W., Oh, C., Lee, J.: Face Detection and Facial Feature Extraction Using Color Snake. In: IEEE International Symposium Industrial Electronics, pp. 457–462 (2002)
16. Shih, Y., Yang, M.: A Collaborative Virtual Environment for Situated Language Learning Using VEC3D. Journal of Educational Technology and Society 11(1), 56–68 (2008)
17. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: IEEE Computer Vision and Pattern Recognition Society Conference, pp. 511–518 (2001)
18. Walczak, K., Wojciechowski, R., Cellary, W.: Dynamic Interactive VR Network Services for Education. In: ACM symposium on Virtual reality Software and Technology, pp. 277–286 (2006)
19. Wang, Q., Ju, S.: A Mixed Classifier Based on Combination of HMM and KNN. In: IEEE 4^{th} International Conference on Natural Computation, vol. 4(18-20), pp. 38–42 (October 2008)
20. Wilson, P., Fernandez, J.: Facial Feature Detection Using Haar Classifiers. Journal of Computing Sciences in Colleges 21(4), 127–133 (2006)
21. Calvi, C., Porta, M., Sacchi, D.: e5Learning, an E-Learning Environment Based on Eye Tracking. In: IEEE International Conference on Advanced Learning Technologies, pp. 376–380 (July 2008)
22. Cristinacce, D., Cootes, T.: Facial Feature Detection Using AdaBoost with Shape Constraints. In: British Machine Vision Conference (2003)
23. Peng, Y., Jin, Y., He, K., Sun, F., Liu, H., Tao., L.: Color Model Based Real-time Face Detection with AdaBoost in Color Image. In: International Conference on Machine Vision, pp. 40–45 (December 2007)

Part II

Health and Medicine Applications

An Enriched Understanding of Why the Environment and Individual Characteristics Are Important in Understanding Technology Utilization in Healthcare: An Evolutionary Psychology Perspective

Chon Abraham[1] and Iris Junglas[2]

[1]College of William and Mary
Chon.Abraham@mason.wm.edu
[2]Accenture Center for High Performance Research Boston, MA
Iris.Junglas@accenture.com

Abstract. In this research, we provide an interdisciplinary perspective intended to enrich our understanding about the individual and the environmental components of technology utilization. We do so by merging the IS utilization literature with evolutionary psychology. Using a qualitative study, we describe two cases of technology utilization in a hospital setting. These two cases were of particular interest as both employ the same technology, involve the same task, occur in the same hospital, and share the same set of users, yet yield different utilization patterns. We provide an integrated model, the Environment-to-Technology-Utilization Model, which is insightful for IS fit research and practice in comprehending varying utilization behavior across work environments.

Keywords: Utilization, fit, technology, task, individual, environment, evolutionary psychology, four drives, healthcare.

1 Introduction

Finding appropriate technology for task performance is an age old problem that humankind has faced since its genesis. The fit between task and technology dictates not only if the tool is used, but also to what extent and in which ways it is used, and if its utilization leads to an increase in task performance or warrants the abandonment of the tool. The decision lies with the individual who, after undergoing a cognitive evaluation process, chooses a tool that will most likely fit the purpose of completing the task [1], [2], [3].

IS research, up to now, has mostly focused on task and technology when evaluating technology utilization, with less attention paid to the task environment. Far less attention has been paid to the individual performer, including what motivates him or her to use a particular tool in that environment. In this research, we provide an interdisciplinary perspective intended to enrich our understanding about the individual

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 141–150, 2011.

and the environmental components of technology utilization. We do so by merging the IS utilization and fit literature, traditionally grounded in social and cognitive psychology, with evolutionary psychology.

Evolutionary psychology purports that human behavior is motivated by mechanisms for survival. These mechanisms evolved over the eons, created as a response to threats from the environment such as food scarcity, threats from other living creatures and dangerous natural elements, or from a lack of companionship or other impediments to social relations. Today, those inherited mechanisms make up our biological repertoire, and while they may not be evident on a daily basis, they are activated more profoundly in situational contexts that resemble those that our ancestors encountered [4], [5], [6], [7].

Healthcare is one such situational context. The importance of "healthcare" can be characterized by its direness due to the imminent threat of debilitating illnesses that would lessen the chances of reproduction and survival. "Healthcare" environments expose a heightened level of emotional distress, not only for the individual whose survival is at stake, but also for his/her family, the caregivers and/or the entire tribe [8]. It is exactly in these environments, reminiscent of ancestral conditions, that the innate set of evolved psychological mechanisms is invoked [6], [7], [9], [10], [11], [12], [13], [14], [15], [16].

Over time, and selected on the basis of survival and reproduction, these mechanisms became innate in the human psychological makeup. In Evolutionary Psychology terms, these mechanisms are labeled as evolved psychological mechanisms (EPM) [7], [11], [17] and represent the initial component within us that motivates behavior [9], [10], [14], [15], [16], [18].

The Four-Drive Model [19], [20] categorizes EPMs based upon what EPMs are designed to appease (i.e., a type of drive or desire that the information processing schema addresses) within the human psyche that drives or motivates behavior. The plethora of EPMs fall into four categories of human drives: the drive to acquire, to bond, to comprehend, and to defend. These drives are "central to a unified understanding of human behavior" [21].

The drive to acquire describes a category of EPMs that motivate a person to seek status, take control, and to retain objects and personal experiences [19]. The drive to bond comprises EPMs that motivate an individual to form social relationships and develop mutual caring commitments with other humans [19]. The drive to comprehend comprises EPMs that push humans to collect information, assess the needs of a situation, examine their environment, and make observations about explanatory ideas and theories to appease curiosity and make sound judgments [19]. And lastly, the drive to defend constitutes a category of EPMs that motivate individuals to defend themselves and their valued accomplishments whenever they perceive them to be endangered [19].

In sum, evolutionary psychology offers a lens for understanding the environment and its interplay with an individual's drives from a different perspective than that offered by the traditional IS literature. The individual, viewed as the end user in technology implementations, is ultimately driven in his motivation to choose a tool (or technology) by a set of innate mechanisms. If these drives, triggered and controlled by the environment, are appeased, pre-determined behavior will follow.

2 Research Design and Methodology

A multi-case study was designed that uses a "within" and "cross-case" analysis. The design is based on Eisenhardt's (1989) model [22] that uses case studies for theory building, as well, as the enhancement of existing relevant literature. Similar to Kirsch (2004) [23], however, this study gleans direction from positivism [24] as the theoretical framework is presented a priori, even though it incorporates tenets and techniques from grounded theory and interpretivism [25], [26], [27], [28]. Thus, we characterize our approach as soft positivism, which applies the existing knowledge of constructs, but allows for concepts to emerge through interpretivism to reveal novel concepts and relationships insightful for both theory building and theory enhancement [23].

2.1 Sites

We identified two cases for study,[1] a Post Anesthesia Care Unit (PACU) and an Ambulatory Care Unit (ACU) that had recently undergone an MICT implementation. In the ACU unit, many different procedures were performed regarding the preparation of patients for surgery, the care of the patient during outpatient surgery or less invasive procedures, and the discharge, or in-hospital transfer of patients upon completion of the procedure. The unit had 30 patient care areas and a waiting room for persons accompanying the patient. PACU was the recovery area post surgery. Upon entry to the unit from the operating room, patients were under anesthesia and were monitored by nursing staff until they regained consciousness. PACU contained 15 patient suites and did not permit visitors.

The same nurses worked both care environments during and after MICT implementation, but there were differing perceptions of the fit of the MICT by nurses working in both units. This is a rare constellation in qualitative research, where we, as researchers, were fortunate to compare two real-life cases that involved the same hospital, the same technology, the same task, the same the set of users, yet with different outcomes. We therefore focused on the subset of nurses to compare the fit of the MICT in each unit.

2.2 Data Collection

Primary techniques for data collection included interviews and direct observations.[2] Data were collected across a sixth month period, beginning four months post the initial implementation. Thus, the data collection was cross-sectional. We conducted 23 audio-recorded interviews across both units averaging 30 minutes for frontline nurses and one hour for directors. All data were transcribed using Atlas.ti®. We used direct observation to assess actual usage behavior, in order to minimize self-report bias. In total, we conducted 35 instances of direct observations of individual nurses

---

[1] The hospital containing the units as cases is a 124-bed is a 25 year-old, not-for-profit hospital. It services more than 110,000 residents in southeast Georgia and was an early adopter of comprehensive electronic medical records, computer physician order entry, and electronic charting for nurse documentation.

[2] Archival documents such as system specifications were reviewed to ensure the MICT had similar purpose and design for each unit.

across units that typically lasted for an entire eight hours shift. Observations were interspersed between interviews to provide a holistic picture of nurses at work that could inspire new or revised interview questions.

3 Findings

As indicated earlier, all three components were invariant, i.e., both cases dealt with the same technology, the same task, and the same individuals. Despite matching similarities on all accounts, utilization differed and ultimate technology abandonment occurred six months post-implementation in ACU while use continued in PACU. It became apparent that individual and environment characteristics akin to those espoused in evolutionary psychology contributed to this divergence.

3.1 Individual Characteristics

Using the Evolutionary Psychology lens for analysis purposes, we will demonstrate that the individual characteristics classified as the drive to acquire, the drive to bond, the drive to comprehend, and the drive to defend impacted utilization behavior.

The ***drive to acquire*** can refer to material objects or status as a desired object or metaphysical state, described by Lawrence and Nohria (2002) [19]. It manifested itself in conjunction with the nurses' concerns of how they were perceived by others. In ACU, aspects of self-presentation emerged, suggesting that nurses were very concerned with how patients and the persons accompanying the patient perceived them, which relates to the referent power of the care giver, as illustrated in the following:

> *ACU Nurse: Doctors tend to get automatic respect from patients with just their title or from just wearing the white coat (**theme: drive to acquire, code: self-presentation associated with mere physical presence**)…I want to have that level of respect…. I need a tool that doesn't detract from that image… I can't let them see me fumbling...I started back to making my notes to put in the chart on the old list or whatever I had to write on…**(memo: lack of willingness to tolerate computer resulting in unintended utilization)**.*

In PACU, nurses were charged with caring for patients who were unconscious for the majority of the time inside the unit, which contributed to minimal verbal communication between the patient and the nurse during that time. Because of the unconscious level of the patient, the drive to acquire status with the patient was less salient in the PACU, even though it existed. In PACU, nurses were less concerned about portraying status in front of the patient as they were concerned about representing themselves as competent amongst fellow medical personnel, as the following excerpt suggests:

> *PACU Nurse: PACU patients are barely conscious while they are in here. So I'm not so worried about them seeing me mess around with the computer if something's not working right...but it really makes me look better with the other nurses on the other units.*

Another dimension of individual characteristics was the theme ***drive to bond*** that refers to our need for social relationships with others. It emerged as very salient in the ACU as nurses deemed it necessary to connect with patients on a personal level. This relationship establishment extends beyond task requirements to a more humanistic display of empathy:

> *ACU Nurse: Our patients are very nervous when they come in here and you do feel for them and want to help ease their nerves (**theme: drive to bond, code: desire to display compassion**)…I lose eye contact with my patient when I try to use the computer in the room. If I can't make eye contact then I can't reassure the patient that I'm here for them and I care…*

At PACU, where patients spent more of their time anesthetized, the drive to bond did not initially appear salient. However, we found out that this drive became of utmost importance once the patients gained consciousness, as the nurses needed to soothe patients' anxiety about their condition. Our field notes read:

> *PACU Nurse: When my kids [affectionately referring to children having undergone surgery] come in after surgery to put tubes in their ears, I get really antsy to just stay by their side (**theme: drive to bond, code: desire to display compassion**).*

The ***drive to comprehend*** impelled nurses to collect information about patients and their ailments. Nurses noted the need to use multiple means of gathering information in order to help them in understanding the patient's conditions. While one might simply explain this as task needs, nurses suggested that sometimes it was simply about satisfying an inner drive, as suggested here:

> *ACU Nurse: I had this patient who use to come for schedule colonoscopies…the nicest sweetest person….this one time he came I noticed he wasn't quite himself…. He kind of snapped at me. …It really bothered me, so later on I read through notes in his chart that indicated he'd had a recent death in this family… information in the chart can help you make sense of what's happening (**theme: drive to comprehend).***

The influence of the ***drive to comprehend*** was most telling in PACU for several reasons. First, patients were most often anesthetized and could not answer questions. Second, nurses deemed the technology, despite its faults, as a means to document insights about the patient's condition upon entry to the unit. And thirdly, the technology allowed nurses access to information about multiple patients' conditions wherever their location, preventing them from having to leave their post with one patient to gather information on another. The following quote underlines this:

> *PACU Nurse: The computer is really the main source you have for information. The patients can't tell you anything and you don't get that much detail from the OR nurses in the hand-off (patients coming from the OR are escorted by the OR nurses into PACU). So I need to use the computer to get as much of that detail to help me make sure I do what's needed for him **(theme: drive to comprehend)**.*

The ***drive to defend***, as part of individual characteristics, manifested itself as the need for the nurses to protect themselves from liabilities. As a result, nurses documented all aspects of the interactions between them, the doctors, and the patients, very thoroughly:

> *ACU Nurse: We are accountable for what we do or don't do right for the patient as well* ***(theme: drive to defend, code: desire to ensure correct accountability).*** *Besides that I take it personal if I know it's something I did or could have prevented that harmed the patient* ***(theme: drive to defend, code: desire to ensure harm prevention).*** *I have to think of myself too in this. If the system doesn't work right it not only puts the patient in jeopardy but me too.*

In PACU, nurses were meticulous in every intervention with the patient and were observed diligently using the information system to validate prescribed medications and regimens as documented in the system. The information system was their primary source of information and served as a means to not only protect themselves, but also the patient from being subjected to harm, as discussed here:

> *PACU Nurse: Our patients can't tell you who they are when they first come in or that they are HIV positive and you need to take extra precautions* ***(theme: drive to defend, code: desire to ensure harm prevention)****….Now [with the MICT] the med list and orders are documented and we can see them in the system as well and know we aren't having to decipher the docs writing and you can document your actions and show proof as to why you did something [used in deposition when medical malpractice is questioned]* ***(theme: drive to defend, code: desire to ensure correct accountability)****.*

Overall, our analysis of the four drives provided insight into the rationale for what salient individual characteristics contribute to technology utilization in these environments, which extends beyond those typically noted in fit literature.

3.2 Environmental Characteristics

As part of the cross-case analysis, and as a part of our evolutionary lens, we identified themes of environmental characteristics which were labeled: *animate inhabitants*, *inanimate objects*, and *atmosphere*. We conceptualize these characteristics and show associations with the four drives that motivate utilization behavior across the units.

As ***animate inhabitants*** we classify the existence of other human beings in the environment. Unlike PACU, ACU was not a completely restricted environment in that family, friends, assigned professional patient advocates, or case managers, may accompany the patient to the unit and, while encouraged to stay in the waiting room, were not restricted to do so. Also in ACU auditory levels and frequency of outbursts, like a patient or companion calling for assistance, patients crying or being consoled prior to procedures because of their anxiety or after having received poor diagnoses, or medical personal calling for assistance to deal with a patient in a dire or critical state, were rather high, making it difficult for certain drives to be appeased.

The theme ***inanimate objects*** denotes the characteristics of the physical environment. Specifically, it describes the existence of objects and artifacts in the environment as well as the architectural layout. The ACU contained single patient rooms or open bays with curtains in between them. In these bays, it was somewhat crowded with multiple beds, tables, and mounted equipment on walls, sinks, and chairs. Maneuverability in these areas with the MICT was often problematic, which not only impeded work, but made it difficult for certain drives to be appeased.

In the PACU, unlike the ACU, patients were not housed throughout the unit, which made it unnecessary for nurses to traverse from one patient room to another with their carts. The physical confines were unobstructed, i.e., there were no dividing walls in PACU, unless a patient was in a highly controlled room due to his condition or level of contagiousness. The open bay setup in PACU enabled MICT maneuverability within and between patient locations. Even though far more monitoring equipment is in use, this equipment is mostly mounted at the head of the bed, and nurses perceived the physical layout as less obstructive. They simply positioned the MICT at the foot of the bed where the display of all vital statistics was easily viewable.

The theme ***atmosphere*** captures the ambiance of the environment or the mood about the state of affairs within it. Dire situations associated with the criticality level of a patient is characteristic of healthcare environments. While PACU and ACU may not have had as dire atmospheres as other units in the hospital, there was still a marked mood of seriousness in both. While, in fact, the criticality level of patients in PACU were graded higher than in ACU, primarily because of the fragile state of patients following surgery, the mood in PACU was much lighter than in ACU. Overhead lights were dimmed in PACU; nurses whispered to each other and spoke in soft, reassuring tones to patients; there were far less jarring noises. Compared to ACU, nurses seemed relaxed and were observed helping each other to operate the MICT; they were also less visibly frustrated when verbally asked questions by other nurses to help them work through technical issues in a cooperative manner. The influence of the drives seems to take precedence in atmospheres where direness, marked by the heightened condition of patients' status, is present.

4 Discussion

Our findings about the individual and environmental characteristics constitute a blend of what is already known and novel characteristics relative to tenets of evolutionary psychology. If we apply traditional social and cognitive conceptualizations for task, technology, and the individual, consistent with, for example, the TPC model [3] or the FITT framework [2], we would have expected that technology for both cases would have been abandoned in both units. Instead, our analysis reveals that utilization behavior and ultimately acceptance can differ even when task, technology, and individuals are invariant across environments.

In ACU the drive to bond and the drive to acquire were quite prevalent, mainly because of the conscious state of the patient. In contrast, in PACU the drive to comprehend and the drive to defend were more observably invoked. Appeasing a salient drive, due to our biological urge, takes precedence. Thus, we propose:

Proposition 1: Characteristics of the environment provoke human drives that an individual will try to appease.

Yet again, the environment, by its virtue of changing its characteristics, can cause new drive(s) to be prevalent and the nurse to defer to a new task after cognitively processing the environment. Performing the task, according to evolutionary psychology, has to be done in order to allay harm that is likely to threaten the sanctity of life, whether for oneself or within one's social setting [7], [44]. In other words, performing the task is motivated by the likelihood that completing the task will appease the underlying drive(s). We suggest:

Proposition 2: As the environmental characteristics change, salient human drives are invoked that initiate the rationalization process for determining the most important task to perform.

Aside from impacting individual drives to be attuned to the most current task, environmental characteristics also invoke drives that influence an individual's technology assessment. In PACU, for example, the MICT was the sole source of information, which appeased the nurses' salient drive to comprehend. Technology can also inhibit the saliency of the drives and challenge their potency. In ACU, for example, the drive to acquire was not supported. We therefore assert the following:

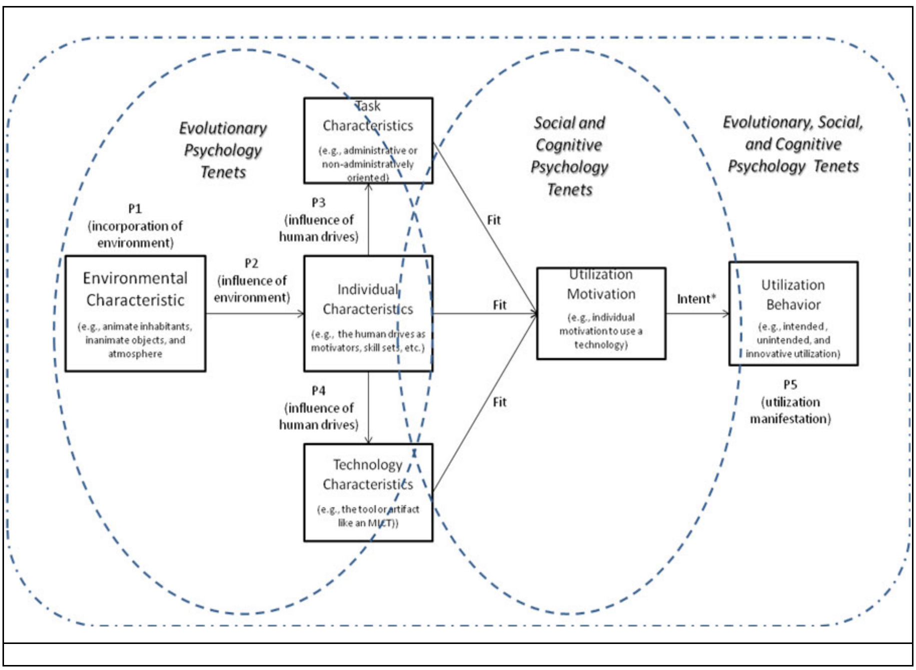


Fig. 1. The Environment-to-Technology-Utilization Model

* Note: Intent is associated with the premise [45] that intentions (in this case the utilization motivation) leads to behavior (in this case the utilization of technology), which is consistent with IS fit literature based on tenets in social psychology.

(Proposition 1), and this invocation of the drives (c) initiates the rationalization process to select the task and technology that will appease the drive (Proposition 2) and (d) motivates utilization behavior via the human drives in their influence on rationalizing the task and technology.

5 Conclusion

Evolutionary psychology is being touted as a unifying discipline, providing a more comprehensive understanding of human behavior that combines elements of social, cognitive, and developmental psychology as well as tenets of neuroscience, as evident in the special issue in the most respected resource *American Psychology* [46]. In summary, the combination of social, cognitive, and evolutionary psychology has the potential to explain more fully human behavior in a wide variety of situations, including many in which humans fund, analyze, design, implement, and use information systems. We came to realize its relevance in this study in the form of the Four-Drive model, which contributes to understanding utilization behavior by providing an enrichment to individual characteristics and environmental characteristics to be incorporated into the IS researchers' theoretical repertoire.

References

1. Malhotra, Y., Galletta, D.F., Kirsch, L.J.: How Endogenous Motivations Influence User Intentions: Beyond the Dichotomy of Extrinsic and Intrinsic User Motivations. J. Management Information Systems 25, 267–299 (2008)
2. Ammenwerth, E., Iller, C., Mahler, C.: IT-Adoption and the Interaction of Task, Technology, and Individuals: A Fit Framework and a Case Study. BMC Medical Informatics and Decision making 6 (2006), `http://www.biomedcentral.com/1472-6947/6/3`
3. Goodhue, D., Thompson, R.L.: Task-technology Fit and Individual Performance. MIS Quarterly 19, 213–236 (1995)
4. Kock, N.: The psychobiological model: towards a new theory of computer-mediated communication based on Darwinian evolution. Organization Science 15, 3327–3348 (2004)
5. Kock, N.: Information systems theorizing based on evolutionary psychology: An interdisciplinary review and theory integration framework. MIS Quarterly 33, 395–418 (2009)
6. Cosmides, L., Tooby, J., Barkow, J.: Introduction: evolutionary psychology and conceptual integration. In: Barkow, J., Cosmides, L., Tooby, J. (eds.) The Adapted Mind, pp. 3–15. Oxford University Press, New York (1992)
7. Buss, D.: Evolutionary psychology: A new paradigm for psychological science. Psychological Inquiry 6, 1–30 (1995)
8. Christensen, C.M., Bohmer, R., Kenagy, J.: Will disruptive innovations cure health care? Harvard Business Review 78, 102–114 (2000)
9. Kenrick, D., Schaller, M., Simpson, J.: Evolution is the new cognition. In: Schaller, M., Simpson, J., Kenrick, D. (eds.) Frontiers of Social Psychology: Evolution and Social Psychology, pp. 1–14. Psychology Press, New York (2006)
10. Pinker, S.: How the Mind Works. Norton, New York (1997)
11. Buss, D.: Evolutionary psychology: A new paradigm for psychological science. Psychological Inquiry 6, 1–30 (1995)
12. Nicholson, N.: How hardwired is human behavior. Harvard Business Review 76, 134–147 (1998)

13. Cosmides, L., Tooby, J.: Evolutionary Psychology and the Emotions. In: Lewis, M., Haviland-Jones, J.M. (eds.) Handbook of Emotions, pp. 91–115. Guilford Press, New York (2000)
14. Wilson, E.O.: Sociobiology: The New Synthesis. Harvard University Press, MA (2000)
15. Wilson, E.O.: On Human Nature. Harvard University Press, Cambridge (2004)
16. Spohn, M.: Organization and Leadership Theory: An Evolutionary, Psychology Perspective. J. of Evolutionary Psychology 26, 97–107 (2005)
17. Duntley, J.D., Buss, D.M.: Evolutionary psychology is a metatheory for psychology. Psychological Inquiry 19, 30–34 (2008)
18. Ruth, W.J.: Evolutionary psychology and rational-emotive theory: Time to open the floodgates. J. of Rational-Emotive & Cognitive-Behavior Therapy 11, 235–246 (1993)
19. Lawrence, P.R., Nohria, N.: Driven: how human nature shapes our choices. Jossey-Bass, San Francisco (2002)
20. Nohria, N., Groysberg, B., Lee, L.: Employee motivation: A powerful new model. Harvard Business Review 86, 78–84 (2008)
21. Lawrence, P. R.: Being Human: A Darwinian Theory of Human Behavior (2007), `http://www.prlawrence.com/beinghumandownload.html`
22. Eisenhardt, K.: Building Theories from Case Study Research. Academy of Management Review 14, 532–550 (1989)
23. Kirsch, L.: Developing Common Systems Globally: The Dynamics of Control. Information Systems Research 15, 374–395 (2004)
24. Yin, R.: Case Study Research: Design and Methods, 2nd edn. Sage Publications, Thousand Oaks (1994)
25. Barrett, M., Walsham, G.: Electronic trading and work transformation in the London Insurance Market. Information Systems Research 10, 1–22 (1999)
26. Boudreau, M.-C., Robey, D.: Enacting Integrated Information Technology: A Human Agency Perspective. Organization Science 6, 3–18 (2005)
27. Golden-Biddle, K., Locke, K.: Composing Qualitative Research. Sage, Thousand Oaks (1997)
28. Madill, A., Jordan, A., Shirley, C.: Objectivity and Reliability in Qualitative Analysis: Realist, Contextualist and Radical Constructionist Epistemologies. British J. of Psychology 91, 1–20 (2000)
29. Shanteau, J.: Competence in Experts: The Role of Task Characteristics. Organizational Behavior and Human Decision Processes 53, 252–266 (1992)
30. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975)
31. Dewsbury, D.: Charles Darwin and psychology at the bicentennial and sesquicentennial: An introduction. American Psychologist 64, 67–74 (2009)

A Real-Time Interactive MIDI Glove for Domicile Stroke Rehabilitation

Nizan Friedman, David Reinkensmeyer, and Mark Bachman

Henry Samueli School of Engineering, University of California, Irvine
Department of Biomedical Engineering
305 Rockwell Engineering Center, Irvine CA USA 92697-2700

Abstract. Stroke is the leading cause of chronic adult disability in Western countries. After several weeks of inpatient physiotherapy, stroke patients are forced to continue unguided and monotonous therapy at home. Consequently, these patients often lose motivation to continue domicile stroke therapy and therefore do not recover to their potential. An interactive real-time MIDI-Glove was developed with the goal of engaging patients in meaningful, entertaining, and motivating domicile therapy. The MIDI-Glove can also provide a quantitative assessment of progress which provides feedback to both patient and therapist. This paper focuses on two developed MIDI-Glove applications. The first is Musiquant, a computer game which allows an individual to play a sample of a song using the glove and to receive a score based on the performance. The second application allows an individual to play along with a song using a variety of different instruments.

Keywords: Interactive real-time MIDI glove, music, stroke rehabilitation, quantitative progress measurement

1 Introduction

Cerebrovascular accident (CVA), or stroke, is the leading cause of chronic adult disability in Western countries with over four million stroke victims currently living in the United States [1]. This number is projected to increase due to an anticipated increase in the geriatric population and decrease in mortality rate of people who have experienced a stroke.

Following the onset of stroke, patients receive several weeks of intensive rehabilitation with the main goal of increasing cognitive and functional abilities [13]. Through intensive and repetitive motion training, patients are able to regain lost function through neural re-organization [8]. Unfortunately, the length of stay at inpatient rehabilitation facilities is usually limited to two weeks and outpatient physical therapy to two times a week [10]. Consequently, patients must continue therapy independently at their home.

Without access to specialized equipment and supervised therapy, the patient is often confined to independent, monotonous tasks. For instance, one task for patients with hand impairment is to passively move flexors and extensors of the affected hand.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 151–158, 2011.

Unfortunately, without the presence of a clinician, this type of task is neither engaging nor does it provide any quantitative measure of improvement—patients therefore lose motivation to perform independent rehabilitation.

A real-time Musical Instrument Digital Interface (MIDI) glove was developed with the goal of providing an innovative approach to at home stroke rehabilitation therapy for the hand. Since over 80% of individuals who experience a stroke suffer hemiparesis, and the most disabling motor deficit is the hand [5], a glove that focuses on hand movement targets a large population. The MIDI-Glove is a low-cost device which both engages patients by providing visual, auditory, and tactile feedback, and provides a quantitative measurement of progress. It focuses on providing calculated highly repetitive motions which are crucial for hand rehabilitation [8].

1.1 Motivation for Auditory Feedback

The integration of auditory feedback with rehabilitation provides distinct advantages over traditional domicile therapy. Music-supported therapy has been shown to lead to significant improvements of motor function in post-stroke patients that can be characterized by better cortical connectivity and improved motor cortex activation [2]. Moreover, neural adaptation due to music performance, even in unskilled individuals, has been seen in integrative auditory sensorimotor areas and therefore is not restricted to cortical motor areas [3]. This suggests that the act of playing an instrument with auditory feedback will allow for better neural plasticity and reorganization.

2 Design Motivation

The glove was designed with the intention of providing meaningful exercise that will help the patient perform everyday tasks. Although the human hand is complex in structure with 27 degrees of freedom [11], a small number of combined joint motions, or synergies, can account for more than 90% of the variance. Using principle component analysis, everyday tasks can be characterized through nine distinct synergies [12]. The figure below depicts the five main movements associated with the designed MIDI-Glove (see Fig. 1). These five movements incorporate 6 of the 9 main synergies and therefore provide meaningful movement that can be applied to everyday activities.

2.1 Generating an Event Using Electrical Contacts

The glove functions by the same working principle as a basic electric switch. When an electrical circuit is open, the electrical contacts are non-conducting. When electrical contacts are closed, the contacts are touching and electricity is free to flow between them. Likewise, when the contact on the thumb of the glove touches each of the five electrical contacts of the fingers, a closed circuit is created (see Fig. 2).

The glove was designed using electrical contacts rather than commonly used pressure or flex sensors in order to force the patients to touch the thumb to one of the corresponding points on the hand. Doing so, the patient practices pinch and grasp motions most pertinent to activities of daily living.

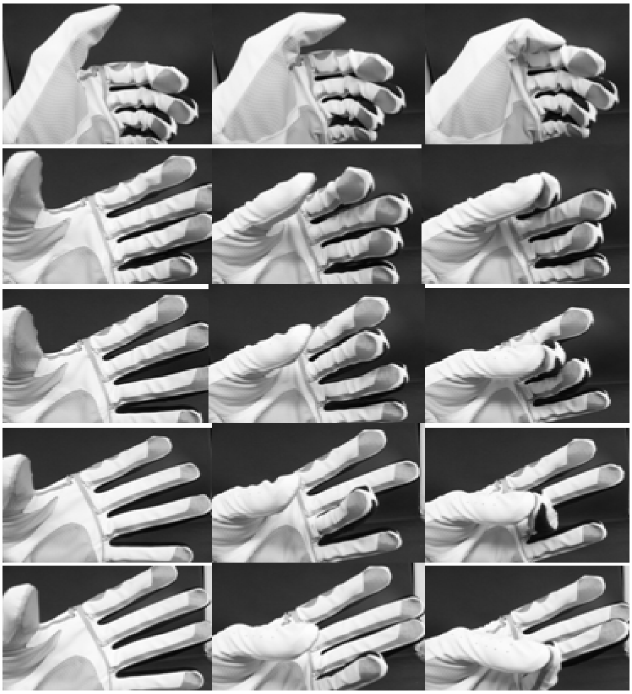


Fig. 1. The five different hand movements associated with the MIDI-Glove. Each row illustrates hand postures related to playing a unique note.

Fig. 2. Five channel real-time MIDI-Glove. When electrical contacts 1-5 contact electrical contact V, a closed electrical circuit is created.

3 Hardware

3.1 USB-MIDI-Controller

Sound from the MIDI-Glove is generated though a MIDI interface, which is the industry standard protocol that enables electronic musical instruments to communicate with computers. By playing sound through MIDI, a virtually endless database of musical instruments is available for the user. MIDI was used in the design in order to allow for large versatility in the both application and creativity of the individual.

The Hale UMC32 USB-MIDI Controller used enables patients to control software applications that support the MIDI protocol through the glove interface. The controller supports up to 32 simultaneous digital or analog inputs allowing the patient to play any or all notes concurrently (see Fig. 3). A sample rate of up to 500Hz can be achieved. The controller is connected to a computer through a Universal Serial Bus (USB) port. The USB-MIDI controller is compatible with all versions of Mac OS X and Windows OS succeeding Windows 2000.

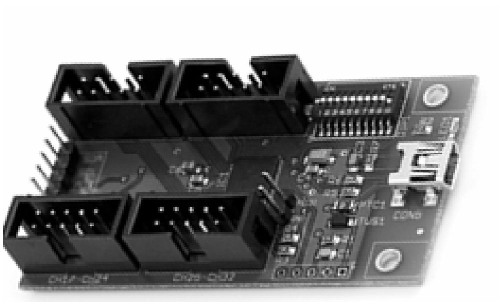


Fig. 3. UMC-32 USB MIDI Controller

3.2 MIDI-Glove

The components of the MIDI-Glove are integrated onto a golf glove. Golf gloves were chosen because they have a Velcro latch on the back which facilitates donning and doffing the glove. Moreover, golf gloves are composed of thin leather material which allows the individual to have superior digit control as opposed to a thicker glove. Numerous vents on the glove allow for increased flexibility and comfort.

The MIDI-Glove has six electrical leads. A nickel and copper plated polyester conductive fabric is used at the fingertips and on the proximal interphalangeal joint on the index finger (see Fig. 2). Conductive fabric was chosen in order to allow for an aesthetic, flexible, conductive material with a large contact surface area. The fabric is highly conductive with a surface resistivity of less than 1Ohm/m^2. Each electrode has a large surface area at the fingertips to force the patient to move the thumb to each of the respective locations on the hand.

3.3 Computer

A Dell Windows XP laptop was used for processing and playback.

4 Software

4.1 MATLAB

The software program MATLAB was used to create the computer program Musiquant. It was also used to calculate the score in Musiquant.

4.2 Reason 5.0

The software program Reason is a streamlined digital audio workstation (DAW) and music sequencer which allows the patient to play and record a variety of musical instruments.

4.3 Hale UMC32 Configuration Utility v1.13

This program allows users to specify which pin of the USB MIDI-controller corresponds to which musical note. This program also allows the user to select whether a digital or analog signal will generate a note. The program also allows the user to configure the sample rate of each channel. For both MIDI-Glove applications a sample rate of 500Hz was chosen.

5 Applications of a Real-Time MIDI-Glove

Using the MIDI-Glove, the patient is currently able to engage in two separate activities.

5.1 Playing Concurrently with a Song

The first application enables a patient to play along to a song using the software program Reason. The patient first selects the desired song. Next, the MIDI-Glove is configured so the five notes are synchronized with the musical key of the song. Since the majority of modern songs, 1960-present, are built on a pentatonic—five note—scale, the individual is able to play along to most songs using this glove. The individual then selects an instrument from the Reason MIDI library and plays along to the song. This explorative approach allows a patient to engage in meaningful finger exercise while being entertained.

Introducing an easy way to play any instrument using this approach may also enable a musician to play an instrument again for the first time using the MIDI-Glove. This may help reduce depression often associated with post-stroke patients who have lost the ability to participate in a musical hobby.

5.2 Playing the Computer Game Musiquant

The second application is a computer game named Musiquant, which quantitatively measures how well a patient plays a song sample. In the computer game, the patient first listens to a sample of a song which concurrently provides visual feedback on which finger must be pressed with respect to each note. Next, the patient plays the song with visual feedback on which note is being activated. Consequently a score of 0-100% is displayed (see Fig. 4).

a.)

b.)

c.)

Fig. 4. Main Graphical User Interface sections of Musiquant computer game. (a) Patient listens to song (b) Patient plays song (c) Patient receives score.

The score is calculated by comparing a patient's performance with a template. The score is dependent on the note correctness (α), the change in onset of note from the template in milliseconds (Δγ), and the change in duration of note from the template in milliseconds (Δβ). These three factors are evaluated for each note.

The score can be expressed as:

$$\text{Score} = 100 - \sum_{n=1}^{K}(1-\alpha)[C_1(\Delta\gamma_n) - C_2(\Delta\beta_n)] - \frac{100\alpha}{K} \quad (1)$$

where: $\alpha = \begin{matrix} 1 & \textit{if wrong note is played} \\ 0 & \textit{if right note is played} \end{matrix}$

C = multiplication factor

K = number of notes played in sample

The multiplication factors C_1 and C_2 are adjusted to determine how significant the change in onset and change in duration of the note is with respect to the score.

This type of application provides both an engaging exercise which motivates the patient to improve, and quantitative feedback. Providing quantitative feedback using Musiquant will help foster a self-competitive and motivating atmosphere. In a recent international randomized clinical trial, stroke inpatients that were given quantitative feedback of their walking speed were able to walk significantly faster than those who were not [6]. Moreover, quantifying progress can also help a therapist obtain a quick yet accurate measure of progress.

6 Future Work

The use of a real-time MIDI glove has yet to be applied to any clinical setting. Longitudinal clinical trials of affected stroke patients provide information on whether the MIDI-Glove is more motivating than traditional at home therapy. Clinical trials will also help identify whether stroke patients benefit from an interactive form of domicile stroke rehabilitation and may also shed insight on whether music coupled with stroke rehabilitation is more beneficial than conventional methods.

From an engineering perspective, the glove can be designed to include other appendages of the body such as the wrist, arm, core, and legs. This may help the patient perform meaningful whole body coordinated movement. Moreover, the glove can be designed to include more notes in order to increase the musical practicality of the glove.

7 Conclusion

Current domicile stroke therapy relies on conventional individual physiotherapy techniques which do not engage the patient in the rehabilitation. The development of an interactive real-time MIDI glove provides a form of meaningful hand therapy that enables the patient to maintain a high level of motivation for hand rehabilitation. Moreover, patients can receive quantitative feedback using the MIDI-Glove which helps foster a desire to improve, and provides physiotherapists with a quick yet accurate measure of progress.

References

1. American Heart Association. Heart Disease and Stroke Statistics—2009 Update. American Heart Association, Dallas (2008)
2. Alternmuller, E., Marco-Pallares, J.: Neural Reorganization Underlies Improvement in Stroke-induced Motor Dysfunction by Music-supported Therapy. The Neruosciences and Music, 395–405 (2008)
3. Bangert, M., Altenmüller, E.: Mapping perception to action in piano practice: a longitudinal DC-EEG study. BMC Neurosci., 26–36 (2003)
4. Chen, J.L., Penhune, V.B., Zattore, R.J.: Listening to musical rhythms recruits motor regions of the brain. Cereb. Cortex, 2844–2854 (2008)

5. Duncan, P.W., Bode, R.K., Min Lai, S., Perera, S.: Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. Arch. Phys. Med. Rehabil., 950–963 (2003)
6. Dobkin, B.H., Plummer-D'Amato, P., Elashoff, R., Lee, J.: SIRROWS Group.: International randomized clinical trial, stroke inpatient rehabilitation with reinforcement of walking speed (SIRROWS), improves outcomes. Neurorehabil. Neural Repair, 235–242 (2010)
7. Kwakkel, G., Kollen, B.J., van der Grond, J., Prevo, A.J.: Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. Stroke, 2181–2186 (2003)
8. Miltner, H.R.W., Bauder, H., et al.: Effects of constraint induced movement therapy on patients with chronic motor deficits after stroke, 586–592 (2004)
9. MusTech, http://mustech.net/2006/09/15/midi-standards-a-brief-history-and-explanation
10. Reinkensmeyer, D., Lum, P., Winters, J.: Emerging technologies for improving access to movement therapy following neurologic injury (2004)
11. Tubiana, R.: Architecture and function of the hand. In: The Hand, pp. 19–93. Saunders, Philadelphia (1981)
12. Thakur, Pramodsingh.: Multidigit Movement Synergies of the Human Hand in an Unconstrained Haptic Exploration Task. The Journal of Neuroscience, 1271–1281 (2008)
13. Walker, M.: Stroke rehabilitation. Br. J. Cardiol. 9, 23–30 (2002)

What Label Design of Ampule for Injection, Do You Want?

Hiroyuki Furukawa

Department of Pharmaceutical Service, Yamaguchi University Hospital, 1-1-1, Minamikogushi, Yamaguchi, 755-8505, Japan
sambista-knz@umin.ac.jp

1 Background

Since "To Err Is Human: Building a Safer Health System (Institute of Medicine publication, USA, 1999)" has published, the interest to accidents which happened in medical institutions has spread to the general public in Japan. And over half number of incident reports in medical institutions is related to medication. The factors which induce medication error are look-alike of the preparations, sound-alike of preparation's trade name, calculation error of dosage and so on.

The healthcare professionals generally recognize the pharmaceutical preparations visually. So, to identify a preparation correctly, the design of the preparation label is important. In the small size preparations like the ampule and the vial for injection. the lots of information are required to write by regulation in the limited narrow space. As a result, the preparation's trade name on the label should be written in small font size, and the medical staff have a potential for the mistaking the preparations and the misunderstanding pharmacological action (therapeutic purpose).

In Japan, some safety measures have been done to the high concentrated potassium preparation for injection which is high risk pharmaceutical agent. One example is to change the label design of high concentrated potassium named "ASPARA KALIUM ™ (Tanabemitsubishi Co. ,Ltd)" and "ASPARAGINSAN KALIUM ™ (Terumo Co. ,Ltd)", which are potassium aspartate preparation for injection widely used in Japan. The part of "ASPARA" is strong product image of nutritional drink "ASPARA (Tanabemitsubishi Co. ,Ltd)" available commercially at the drug stores in Japan, and not few health care professionals misunderstand the pharmacological action of "ASPARA KALIUM ™" .

Therefore, we have made the proposal of changing the font size of the "ASPARA " part and "ASPARAGINSAN" part (smaller) and the "KALIUM" part (larger) to the pharmaceutical company for preventing medication errors of the healthcare providers. And about 2 years after, our proposal has accepted and released the new printed preparation in Nov., 2008 and Mar., 2009, respectively (Fig. 1,2).

2 Purpose

After the preparation of new label design available in the market, we tried to compare the understanding the therapeutic purpose before and after the label design change with the healthcare providers (physician, nurse and pharmacist) and their students as the survey subjects and to evaluate the effect of the label design change.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 159–166, 2011.

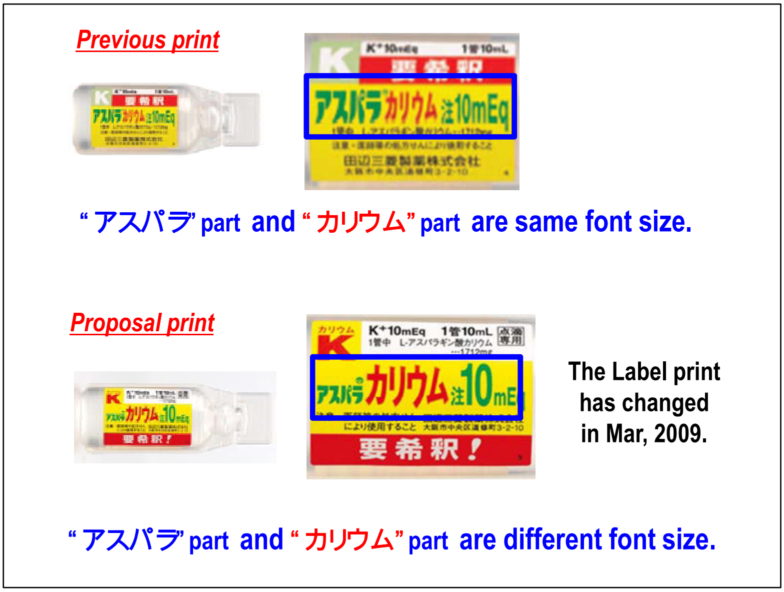


Fig. 1. Previous and new preparation label design of “ASPARA Kalium™ (potassium aspartate)

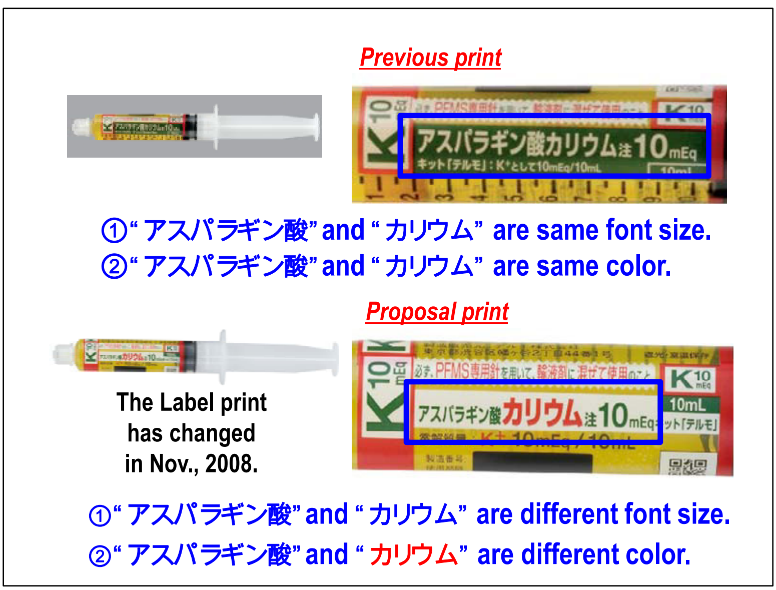


Fig. 2. Previous and new preparation label design of “ASPARAGINSAN Kalium™ (potassium aspartate)

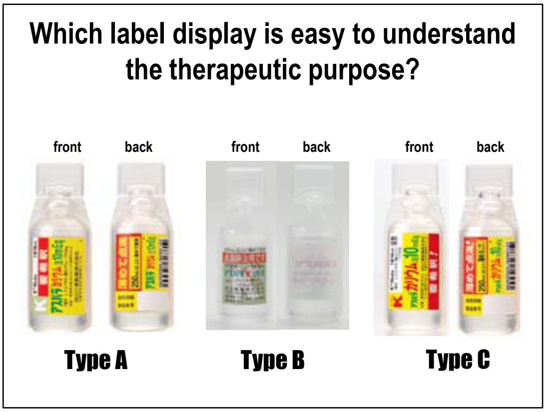


Fig. 3. Survey sheet of ASPARA Kalium ™ (potassium aspartate) for the healthcare providers

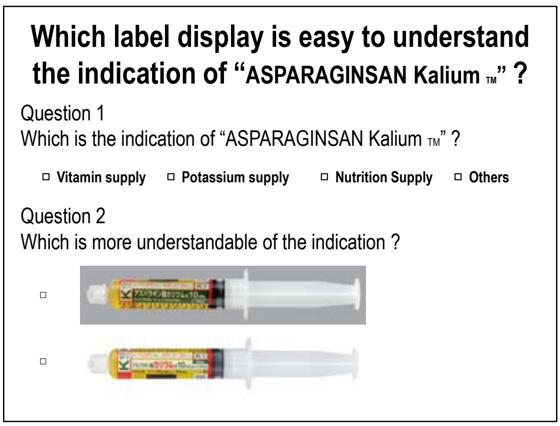


Fig. 4. Survey sheet of ASPARAGINSAN Kalium ™ (potassium aspartate) for nurses and pharmacists

3 Methods

The survey sheets of the picture image of preparation (Fig. 3-6) are showed to the subjects, and the answers were collected on a voluntary basis. The results are analyzed statistically.

4 Results

The survey sheets were collected from physician (n=64), nurses(n=413) and pharmacist (n=328)., and 3 0f 64 physicians, 33 of 413 nurses and 1 of 328 pharmacists did not understand the therapeutic purpose of "ASPARA Kalium ™". (Fig. 7).

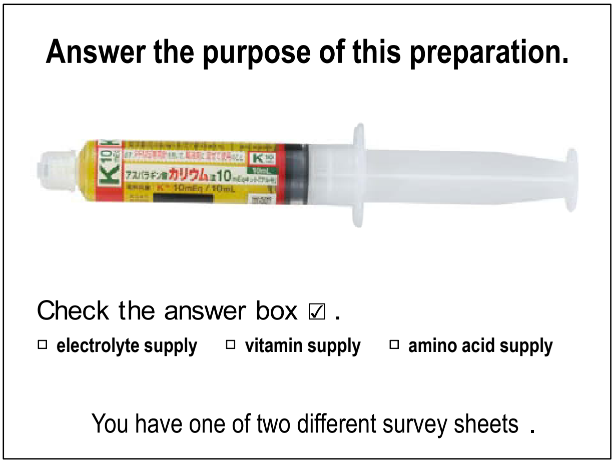


Fig. 5. Survey sheet of ASPARAGINSAN Kalium™ (potassium aspartate)'s new label design for the medical students

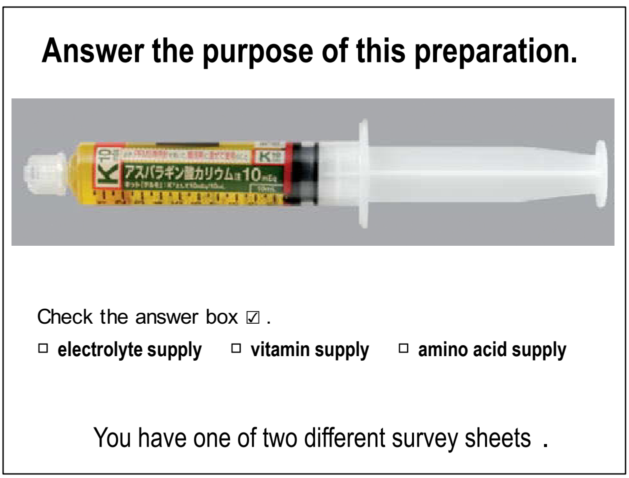


Fig. 6. Survey sheet of ASPARAGINSAN Kalium™ (potassium aspartate)'s previous label design for the medical students

As for "ASPARA Kalium™", our proposal design "Type C" was the most understandable in all three subject groups which are 61 physicians, 380 nurses and 327 pharmacists except the respondents of the therapeutic purpose misunderstanding (fig. 8). And our proposal design was more understandable in the medical students (n=78), but not significant statistically (fig. 9).

As for "ASPARAGINSAN Kalium™", our proposal design was more understandable than the previous design in both the pharmacists (n = 158) .and the nurses (n = 159), and also in the medical students (n=66), statistically significantly ($p<0.05$, chi-square test) (fig. 10, 11).

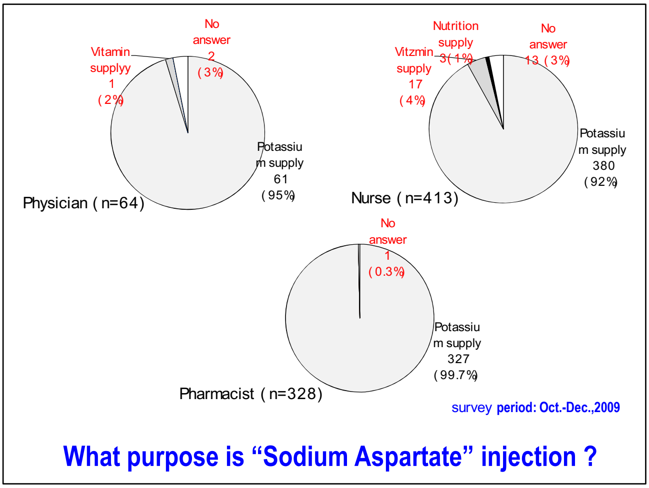


Fig. 7. Understanding of the therapeutic purpose of "ASPARA Kalium ™". (subjects= 64 physicians, 413 nurses and 328 pharmacists. survey period: Oct.-Dec.,2009).

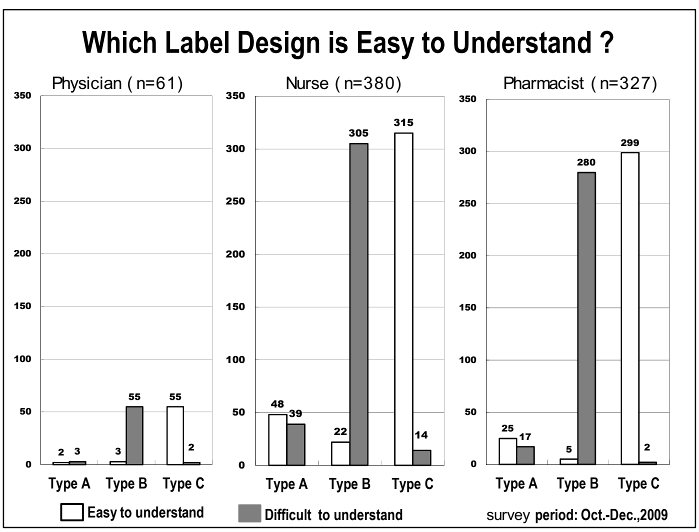


Fig. 8. Comparison of understanding of the therapeutic purpose of "ASPARA Kalium ™" in three label designs. (subjects = 61 physicians, 380 nurses and 327 pharmacists. survey period: Oct.-Dec.,2009).

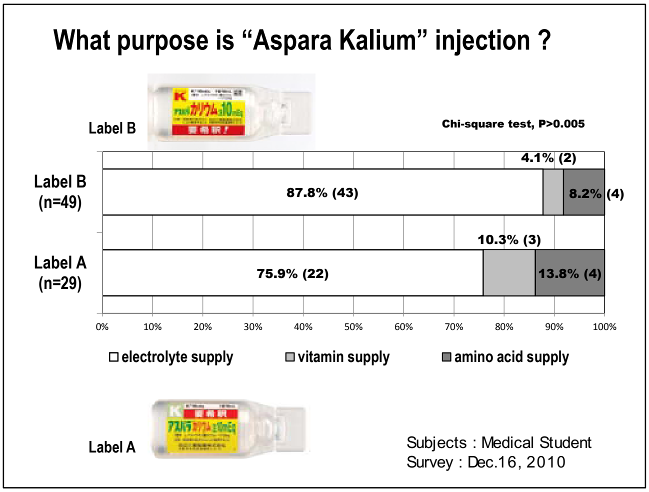


Fig. 9. Comparison of understanding of the therapeutic purpose of "ASPARA Kalium TM" in two label designs. (subjects = 78 medical students. survey period: Dec.,2010).

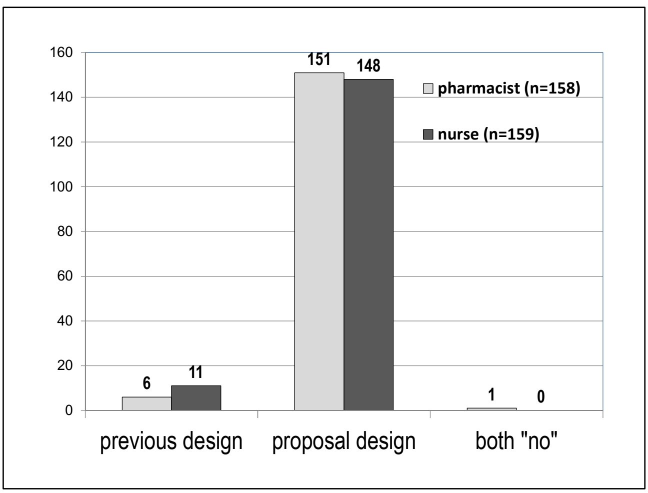


Fig. 10. Comparison of understanding of the therapeutic purpose of "ASPARAGINSAN Kalium TM" in two label designs. (subjects= 158 pharmacists and 159 nurses. survey period: Jan.,2010).

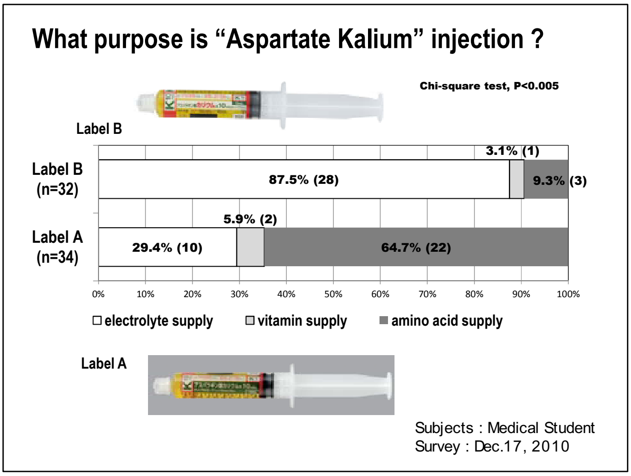


Fig. 11. Comparison of understanding of the therapeutic purpose of "ASPARAGINSAN Kalium ™" in two label designs. (subjects= 66 medical students. survey period: Dec.,2010).

5 Discussion

The injectable preparation of high concentration potassium is high risk when administered rapid intravenously, because it induces cardiac arrest. So the several improvements of the preparation form and design have been tried over the past decade in Japan (fig. 12).

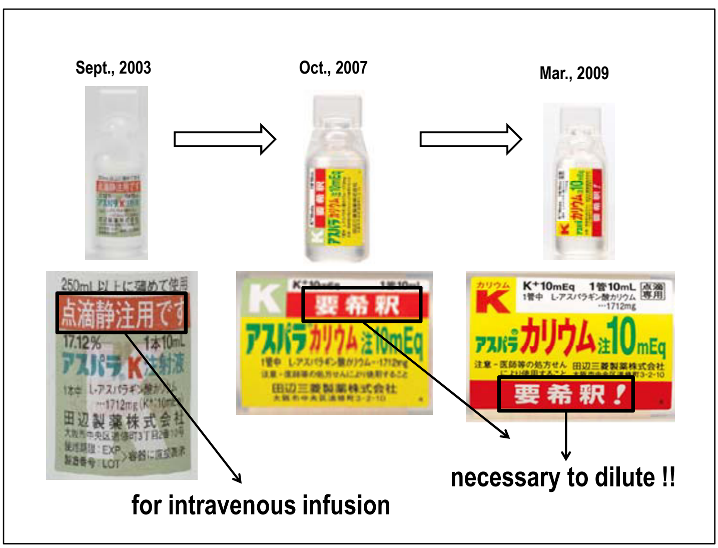


Fig. 12. Change of the label design of "ASPARA Kalium ™"

But therapeutic purpose of some preparations like "ASPARA Kalium TM" and "ASPARAGINSAN Kalium TM" were misunderstood in the healthcare providers and medical students. Especially, a Japanese translation of "aspartate" is "ASPARAGINSAN", so many healthcare providers understand as amino acid supply preparation. And many medical students misunderstand the therapeutic purpose of "ASPARAGINSAN Kalium TM".

So we were proposing the change of the preparation label design, which were the font size and color change of two parts ("ASPARA" or "ASPARAGINSAN" and "Kakium") to the pharmaceutical companies (Fig. 1,2). Our proposes have been accepted and the new designed preparations hav released in Nov., 2008 and Mar., 2009,respectively.

After the change of the preparation design, we tried to evaluate the change of understanding of therapeutic purpose in the healthcare providers. And the all results of survey showed understanding of therapeutic purpose increased by changing the label design. The change of the label design in "ASPARA Kalium TM" and "ASPARAGINSAN Kalium TM" is expected to decrease human error.

6 Conclusion

Changing font size and color of two parts ("ASPARA" or "ASPARAGINSAN" and "Kakium") increased understanding of the therapeutic purpose. And visual improvement in the label design of pharmaceutical preparation is expected to decrease human error.

Acknowledgment. I am very grateful to Fumito Tsuchiya, Professor of Department of Pharmaceutical Science, International University of Health and Welfare, for his advice.

The Design of an Interactive Stroke Rehabilitation Gaming System

Linda Harley[1], Scott Robertson[2], Maribeth Gandy[2], Simeon Harbert[1], and Douglas Britton[1]

[1] Georgia Tech Research Institute,
640 Strong Street, Atlanta GA 30318, United States
{Linda.Harley,Sim.Harbert,Doug.Britton}@gtri.gatech.edu
[2] Interactive Media Technology Center, Georgia Institute of Technology,
85 5th Street NW, Atlanta GA 30308, United States
{scott,maribeth}@imtc.gatech.edu

Abstract. There is a compelling need to create an alternative and affordable home based therapy system founded on sound rehabilitative principles, that is readily available, engaging and motivational, and can be remotely monitored by therapists. In the past two years, stroke related medical costs have increased 20%, while the number of clinical treatment sessions have declined. The purpose of this study was to develop an affordable interactive stroke rehabilitation gaming experience based on therapeutic fundamentals that can easily be used in the clinical setting or the home environment.

Keywords: rehabilitation, gaming, health, upper extremity.

1 Introduction

Cerebrovascular accidents (strokes) are a major cause of disability and the third leading cause of death in the United States, with 780,000 new and recurrent strokes occurring every year. In the past two years, stroke-related medical costs have increased 20% [1] while the numbers of clinical treatment sessions have declined. Accordingly, there is a need to create a home-based stroke rehabilitation system that is affordable, predicated upon sound rehabilitative fundamentals while remaining compatible with interests and activities that promote compliance. Merging existing therapy methods with an evolving notion of "gaming" from the realm of entertainment toward more serious endeavors has the potential to produce a unique rehabilitation learning experience for stroke survivors.

1.1 Approaches to Stroke Rehabilitation

Traditional customary care involves time intensive treatment from an occupational therapist. The rehabilitation strategy depends largely on the goals of the stroke survivor and their current capabilities with regards to range of motion, strength and functional

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 167–173, 2011.

status. This form of rehabilitation often involves performing repetitive movements, which have been shown to induce motor learning and therefore improve the movement behaviour of the stroke survivor [2-3]. Several alternative rehabilitation methods have been proposed. Constrained Induced Movement Therapy (CIMT) involves having the stroke survivor wear a mitten on the healthy arm for long periods of time, thereby restricting movement of the healthy arm and forcing the use of the affected arm [4]. CIMT is only useful for a sub-group of the stroke population because of the mental and psychological requirements in having self-discipline to stick with the program. Another alternative looks at ways to accomplish rehabilitation by using robotics. The notion here being that robots can perform and measure repetitive movement without growing tired, thereby allowing the occupational therapist the ability to work with multiple stroke survivors at once. The draw back to this is that robotic solutions are not cost effective for small business nor is it feasible for the home environment. Robotics also requires some level of skill to operate the robot appropriately. Current paradigms in robotic motor learning involve simple reaching pattern tasks that are not very engaging or motivational, however have been shown to significantly improve the movement behaviour of a stroke survivor [5-7].

Studies have shown that the Wii can have a beneficial health impact on stroke survivors [8-9]. However, current consumer video games for consoles, such as the Wii, require precise dexterous movements and are designed around completing game activities under extreme time pressure; two features that are inappropriate for stroke rehabilitation therapy. Several systems have been designed specifically for stroke intervention, such as the TheraJoy, T-WREX or the RUPERT [10-13]. However, such systems assume that the stroke survivor is able to both grasp and manipulate a joystick or don an exoskeletal orthosis/robot. These systems also do not contain interactive motivational gaming experiences to keep patients engaged, and are too expensive and complex for home use.

1.2 Benefits of gaming

There is mounting evidence suggesting that playing video games produces gains in certain physical and cognitive abilities. Previous studies using video games as the primary intervention have found improvements in reaction time, short-term and long-term memory, attention [14], and spatial ability. While the majority of this research has been conducted with adolescent and young adults, there are a few studies that have also explored video games as an intervention for stroke survivors. Gains in prose recall were found in a sample of stroke patients with memory impairment after interaction with memory games [15]. Other investigators have speculated that playing videogames may actually stimulate neurons common to the visuomotor pathways that lead to improved reaction time [16]. While prior investigations have explored the design of gaming systems focused on memory and prose recall for the stroke population, few efforts have attempted to encode specific rehabilitative movements as part of an interactive gaming experience.

1.3 Purpose

The primary goal of this research effort was to create a compelling and fun game/experience that users would enjoy playing and that would also improve the mobility and dexterity of their stroke-affected arm. Through the intersection of Rehabilitation Medicine, Computer Science and Engineering, a proof-of-concept investigation of an innovative home-based stroke rehabilitation technology was developed that encoded a computer video game experience with a limited number of specific rehabilitation movements that were based on sound rehabilitative principles.

2 System Hardware Architecture

Stroke survivors are often not able to grasp and manipulate objects with their affected hand. Therefore careful consideration was given to the type of input device and its attachment to the arm.

2.1 Input Devices

The stroke rehabilitation game utilizes a Wii Remote motion tracking controller from Nintendo's Wii game console. The Wii Remote includes an infrared (IR) camera that can be used to track the position and orientation of a set of IR beacons. A prototype version of the stroke rehab game uses two IR beacons rigidly attached to the player's stroke affected arm and tracked by a ceiling mounted Wii Remote's IR camera. This configuration allows accurate 2D position tracking of the player's forearm.

The raw, 2D forearm tracking data is filtered using a spring-damper function, the output of which is used as the end-effector position of a simple inverse-kinematic (IK) model of a human arm, including a shoulder joint and an elbow joint. The internal IK arm modeling allows reasonably accurate arm motion tracking in a horizontal plane extending from the player's chest forward and perpendicular to the floor.

2.2 System Display

The development and testing of the system was done on a personal computer. This allowed us to create a system that has a more powerful Central Processing Unit (CPU) and Graphics Processing Unit (GPU) than the Nintendo Wii console. A future design goal for the game display would be to incorporate a low cost portable console that could connect directly to a television.

3 System Software Architecture

The prototype stroke rehabilitation game was developed in the Unity 3D game engine. Unity provides a rapid game prototyping and development environment and is based on the Mono .NET software framework. Unity supports cross-platform game development and supports most of the Mono/.NET framework capabilities, including Unix sockets inter-application communication. Unity's Mono/.NET architecture was used to develop a custom Unix sockets client/server interface to communicate with a Nintendo Wii remote which is used as the primary game input device. A separate Wii remote server application was developed to communicate with and receive Wii remote telemetry data such as IR beacon spatial locations. The Wii remote server communicates with the Unity game via a Mono/.Net sockets client object inside the game. During game play, the Wii remote continuously reports IR position tracking data to the game application. Wii remote accelerometer data and button states are also reported and this data may be utilized in future versions of the stroke rehabilitation game.

3.1 System Operation

The stroke rehabilitation game includes most of Peggle's game mechanics and behaviors, but is augmented to facilitate stroke-affected arm motions and exercises. Whereas Peggle uses a mouse or keyboard to aim and fire balls, the stroke rehab game utilizes a Wii Remote motion tracking controller from Nintendo's Wii game console.

3.2 System Content

A number of video game styles and genres were considered as prototypes for a suitable rehabilitation game. Ultimately, the popular casual game Peggle was chosen as a model for the stroke rehabilitation game (Figure 1). This greatly simplified and sped up the graphical and game play design phases and allowed more time and resources to be spent on augmenting and adapting the game to support stroke rehabilitation in the form of encouraging specific arm motions and exercises.

Peggle itself is loosely based on older, mechanical games such as pinball and pachinko (a Japanese version of pinball). Peggle consists of a grid of pegs and other targets, all of which are assigned point values, and allows the user to sequentially fire balls, much as with pinball and pachinko machines, onto the play field with the goal of hitting, and thereby destroying, as many pegs as possible. The player's score accumulates as pegs are hit and destroyed. In the original Peggle game, the player can aim a ball-firing canon. Skill and strategy are required to clear the peg field efficiently, but chance also comes into play as a physics engine controls the ball motion and ball behavior is difficult to predict after the first few ball-peg collision interactions.

A variety of game behaviors can be triggered and controlled via large and small arm motions in the tracking plane previously described. The game control motions were chosen to mimic currently used stroke rehabilitation arm exercises and include precision aiming of the ball canon, positioning of a catch bucket at the bottom of the peg field (used to catch balls as they fall, and score extra points and turns), increasing a point-multiplier "power" value prior to firing a ball. Each such control regime can be tailored, even during a game session, to encourage specific arm motions and exercises such as maximizing affected arm range of motion, training arm motions for precision, training arm motions for stability and minimizing tremor.

4 System Design and Usability

Stroke survivors tend to have limited range of motion, slow reaction time and impaired cognitive skills. Therefore this game had to be based somewhat on chance, yet still be physically challenging and motivational for the individual to continue.

Fig. 1. The prototype of the rehabilitation gaming system

4.1 Design Criteria

Knowledge gathering exercises were conducted with two experts in the field of stroke rehabilitation to ensure that the type of arm motions and exercises to be encoded into the game could lead to improvements in stroke survivors. In designing the game, a number of design criteria were developed which would facilitate stroke rehabilitation movements and exercises yet still be fun and compelling. The game rules and mechanics should be simple and easily understood with the right balance of skill, chance and strategy: enough skill required to facilitate rehabilitating motions and enough chance and strategy required to create compelling game play and motivate the player to play more often and for longer periods of time.

To create an effective and enjoyable game, this initiative followed an iterative game design process. A small set of rehabilitative movements were to be encoded in the game and used for rehabilitation (informed by metrics such as the WolfMotor Function Test [17].

The game design required additional considerations due to the uniqueness and limitations of the target users and to the functional requirements for successful physical rehabilitation. In particular, to achieve true physical gains the players would need to spend significant time in the environment; thus the intervention must be engaging and playable for adults of any age. A game that is boring and tedious or an interface that is too physically strenuous would be unacceptable. A game mechanic that is too easily mastered or that is too difficult can cause player frustration and nullify benefits of the system.

Designing an environment that leads to a feeling of engagement, immersion and "oneness" with the system is more of an art than a science. However, extensive research exists in the area of computer interfaces for older adults, gaming for non-traditional populations, and general game design, including the new area of "serious" game design [18]

4.2 Usability Testing

Due to limited budget and timeline a case study was conducted where a recent stroke survivor volunteered to evaluate the system. The stroke survivor had persistent hemiparesis that lead to impaired upper extremity function but had no severe weakness or sensory impairment of the upper extremity. The individual was able to understand and follow instructions and had a demonstrated interest in technology and games. Upon review of the gaming environment the individual had found that the use of IR motion tracking via Wii remote enabled him to successfully play the Peggle-based game. The individual confirmed that the initial chosen control regime for the rehabilitation game, tracking stroke-effected arm motion in a 2D, horizontal plane, closely approximated traditional stroke rehabilitation exercises and would therefore likely also prove beneficial. Further, the individual stated that the prototype Peggle-based game was fun and conducive to extended use, a primary goal of this research.

5 Conclusion and Future Work

This paper has described the features, design, implementation and operation of the stroke rehabilitative gaming system and preliminary findings from a case-study.

The prototype stroke rehabilitation game currently supports arm-tracking game control in 2D using only the Wii Remote's IR camera tracking ability. Future development will include also utilizing the Wii Remote's thre-axis accelerometers and the three-axis gyroscopes of the MotionPlus extension device. Combining all three sensor types, and attaching the Wii Remote to the player's forearm instead of the IR beacons, should allow full six-degrees-of-freedom (6DOF) tracking of the player's arm. Full, 6 DOF tracking of the player's arm will allow a much larger range of trackable arm motions as well as many more exercises and motions to be encouraged by stroke rehab game play.

Acknowledgments. The authors would like to thank Dr. Steven Wolf and Dr. Sarah Blanton for their assistance in providing the expertise on stroke rehabilitation.

References

1. Pricewaterhouse Coopers Health Research Institute: Behind the Numbers. In: Mdeical cost trends for 2009. Report published (June 2008), `http://pwchealth.com/cgi-local/hcregister.cgi?link=reg/numbers2009.pdf` (last accessed April 17, 2009)
2. Krebs, H.I., Volpe, B., Hogan, N.: A working model of stroke recovery from rehabilitation robotics practitioners. Journal of Neuro. Engineering and Rehabilitation 6(6) (2009)

3. Boyd, L.A., Vidoni, E.D., Wessel, B.D.: Motor learning after stroke: is skill acquisition a prerequisite for contralesional neuroplastic change? Neuroscience Letters 482, 21–25 (2010)
4. Wolf, S.L., Winsetin, C.J., Miller, J.P., Taub, E., Uswatte, G., Morris, D., Giuliani, C., Light, K.E., Nichols-Larsen, D.: Effect of Constraint-Induced Movement Therapy on upper extremity function 3 to 9 months after stroke: The EXCITE randomized clinical trial. JAMA 295, 2095–2104 (2006)
5. Lum, P.S., Burgar, C.G., Shor, P.C., Majmundar, M., Van Der Loos, M.: Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. Archives of Physical Medicine and Rehabilitation 83, 952–959 (2002)
6. Volpe, B.T., Krebs, H.I., Hogan, N., Edelstein, L., Diels, C., Aisen, M.: A novel approach to stroke rehabilitation – Robot-aided sensorimotor stimulation. Neurology 54, 1938–1944 (2000)
7. Volpe, B.T., Krebs, H.I., Hogan, N., Edelsteinn, L., Diels, C.M., Aisen, M.L.: Robot training enhanced motor outcome in patients with stroke maintained over 3 years. Neurology 53, 1874–1876 (1999)
8. Yong Joo, L., Soon Yin, T., Xu, D., Thia, E., Pei Fen, C., Kuah, C.W., Kong, K.H.: A feasibility study using interactive commercial off-the-shelf computer gaming in upper limb rehabilitation in patients after stroke. Journal of Rehabilitation Medicine 42, 437–441 (2010)
9. Saposnik, G., Teasell, R., Mamdani, M., Hall, J., McIlroy, W., Cheung, D., Thorpe, K.E., Cohen, L.G., Bayley, M.: Effectiveness of virtual reality using Wii gaming technology in stroke rehabilitation: a pilot randomized clinical trial and proof of principle. Stroke 41, 1477–1484 (2010)
10. Johnson, L.M., Winters, J.M.: Enhanced TheraJoy technology for use in upper-extreimty stroke rehabilitation. In: Proceedings for the 26th Annual International Conference of the IEEE EMBS, San Francisco, USA (2004)
11. Reinkensmeyer, D.J., Pang, C.T., Nessler, J.A., Painter, C.C.: Web-based telerehabilitation for the upper extremity after stroke. IEEE transactions on Neural Systems and Rehabilitation Engineering 10(20), 102–108 (2002)
12. Sugar, T.G., He, J., Koeneman, E.J., Koeneman, J.B., Herman, R., Huang, H., Schultz, R.S., Herring, D.E., Wanberg, J., Balasubramanian, S., Swenson, P., Ward, J.A.: Design and control of RUPERT: a device for robotic upper extremity repetitive therapy. IEEE Transactions of Neural Systems and Rehabilitation Engineering 15(3), 336–346 (2007)
13. Willmann, R.D., Lanfermann, G., Saini, P., Timmermans, A., te Vrugt, J., Winter, S.: Home stroke rehabilitation for the upper limbs. In: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, France (2007)
14. Green, C.S., Bavelier, D.: Action-video-game experience alters the spatial resolution of vision. Psychological Science 18(1), 88–94 (2007)
15. Towle, D., Edmans, J.A., Lincoln, N.B.: Use of computer-presented games with memory-impaired stroke patients. Clinical Rehabilitation 2(4), 303–307 (1988)
16. Dustman, R.E., Emmerson, R.Y., Steinhaus, L.A., Shearer, D.E., Dustman, T.J.: The effects of videogame playing on neuropsychological performance of elderly individuals. Journal of Gerontology 47, 168–171 (1992)
17. Wolf, S.L., Catlin, P.A., Ellis, M., Archer, A.L., Morgan, B., Piacentino, A.: Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke. Stroke 32, 1635–1639 (2001)
18. Kelly, H., Howell, K., Glinert, E., Holding, L., Swain, C., Burrowbridge, A., Roper, M.: How to build serious games. Communications of the ACM 50(7), 58–64 (2007)

Therapeutic Category Improvement Method Based on the Words Appearing in Effect-Efficacy Description

Hirotsugu Ishida[1], Keita Nabeta[1], Masaomi Kimura[2], Michiko Ohkura[2], and Fumito Tsuchiya[3]

[1] Graduate School of Engineering, Shibaura Institute of Technology, 3-7-5 Toyosu, Koto Ward, Tokyo 135-8548 Japan
[2] Shibaura Institute of Technology, 3-7-5 Toyosu, Koto Ward, Tokyo 135-8548 Japan
[3] International University of Health and Welfare, 2600-1 Kitakanemaru, Ohtawara City, Tochigi, Japan
{m110013,m709102}@shibaura-it.ac.jp, {masaomi,ohkura}@sic.shibaura-it.ac.jp, ftsuchiya@iuhw.ac.jp

Abstract. Medical drugs have various efficacies, and are classified focusing on their purpose of use. In Japan, the Ministry of Internal Affairs and Communications gives Japan standard commodity classification (JSCC) numbers to drugs. Therapeutic category numbers are decided based on three digit numbers after the head digits “87”. Although the current JSCC numbers are determined based on the revised document “Japan standard commodity classification” compiled in 1990, they have not been revised for 20 years. As a result, when drugs are categorized based on this categorizing system, some drugs are not applicable to any category. As the result, the drugs have been categorized as “other categories” such as “drug for other allergy” or “drug for other cardiovascular disease.” The number of such drugs is increasing. However, since it is conceivable that drugs having similar efficacy are often included in other categories, it is necessary that such drugs are classified independently from the “other categories.” Therefore, in this study, we analyzed drugs information categorized as “drugs for other cardiovascular disease,” and proposed a method of classifying these drugs by using clustering.

Keywords: Medical Safety, Therapeutic Category, Clustering.

1 Introduction

Medical drugs have various efficacies, and they are classified focusing on the purpose of use. In Japan, the Ministry of Internal Affairs and Communications [1] gives Japan standard commodity classification (JSCC) numbers to drugs, which are composed of 5 or 6 digits. Ethical and proprietary drugs are given JSCC numbers whose head two digits are “87,” then therapeutic category numbers are decided based on three digits of numbers after “87.” Therapeutic category numbers have a hierarchical structure [2], and the third digit following the two digits “87” expresses the action part of the body

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 174–181, 2011.

or the efficacy. The fourth digit expresses an ingredient or the action part of the body, and the fifth digit expresses the use. For example in the case of "87214," the third digit "2" expresses "drug for individual organic systems," the fourth digit "1" expresses "drug for cardiovascular disease," and the fifth digit "4" expresses "a hypotensive drug."

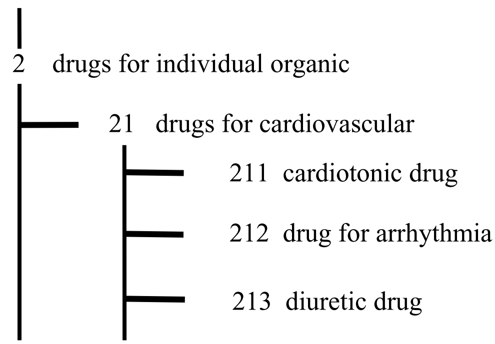


Fig. 1. Part of the hierarchical structure of the therapeutic category numbers

Although the current JSCC numbers are determined based on the revised document "Japan standard commodity classification" compiled by the Ministry of Internal Affairs and Communications, they have not been revised for 20 years. As a result, when drugs are categorized based on this categorizing method, some drugs are not applicable to any category. Therefore, the drugs have been categorized as the "other category" such as "drug for other allergy" or "drug for other cardiovascular disease." The number of such drugs is increasing.

However, since it is conceivable that drugs having similar efficacy are often included in other categories, it is necessary that such drugs are classified independently from the "other categories." Therefore, in this study, we analyzed drugs information categorized as "drugs for other cardiovascular disease," and proposed a method of classifying these drugs by using clustering.

2 Target Data

We obtained 166 drug package inserts formatted with Standard General Markup Language (SGML) whose therapeutic category number is "219," which corresponds to "drug for other cardiovascular disease," from the Pharmaceuticals and Medical Devices Agency (PMDA) [3]. Package inserts are exclusive legal documents used to describe detailed information for each drug, including the composition, efficacy, dosage and cautions. As plural drugs are described in one package insert for every standard, we targeted 235 drugs.

3 The Method Based on Therapeutic Category Names

Firstly, we obtain the therapeutic category names from each drug. Therapeutic category names are defined as "if you can express definitely the efficacy or the character of such drug, you mention them and avoid expressions that may invite misunderstanding by the user" in the notice [4] given in 1997. As such therapeutic category names are decided by each pharmaceutical company, there is no standardization. As a result, there are therapeutic category names such as "circulatory disease improvement drug" based on the concrete efficacy of the drugs and "prostaglandin E1 drug" based on ingredient names of the drugs. Although we can find the efficacy of the drugs directly based on the concrete efficacy as in the former case, we cannot find the efficacy of the drugs directly based on only ingredient names as in the latter case. In addition, some drugs are not given therapeutic category names. Therefore, we cannot classify drugs having similar efficacy based on therapeutic category names.

4 Proposed Method

As mentioned above, since we cannot classify drugs based on therapeutic category names, we classify drugs based on information described on the part of "effect-efficacy" in the package inserts. We generate networks based on the words included in the part of "effect-efficacy" and aggregate drugs with similar efficacy by applying a clustering method to the networks.

4.1 Extraction of Nouns

We divide the statements of "effect-efficacy" into clauses with "CaboCha," which is the Japanese dependency analyzer, and extract nouns included in these clauses. In addition, if we extract compound words that plural nouns connect like "頭部外傷後遺症（head injury aftereffects)," we divide them according to their meanings like "頭部(head)," "外傷(injury)," "後遺症(aftereffect)" and extract these nouns.

Then we count up the number of nouns that are extracted by the method for every drug. Nouns such as "改善(improvement)," "下記(follows)," "もの(thing)" and so forth appear for many drugs. Since we expect that they are not related with the effect directly, we exclude such nouns from the target of analysis. In addition, we exclude nouns that are provided by dividing the compound words and not related with the effect directly. For example, we exclude "慢性(chronic)," which was provided by dividing "慢性腎不全(chronic renal insufficiency)" and leave "腎不全(renal insufficiency)."

4.2 Generation of Networks

We make a connection matrix with the remaining nouns and the product names as a network with them as Fig. 2. A row of the connection matrix expresses what kind of nouns appear in statements of "effect-efficacy" of the package insert of a drug. If a noun appears in a drug statement, the element falling under the noun of the row of the drug is 1. If the noun does not appear in the drug statement, it is 0. Then we make an adjacency matrix between the product names based on the connection matrix as Fig. 3. If drugs connect through common nouns in the connection matrix, we set up an edge between the drugs in the adjacency matrix. Therefore, the adjacency matrix expresses the network of the product name of drugs.

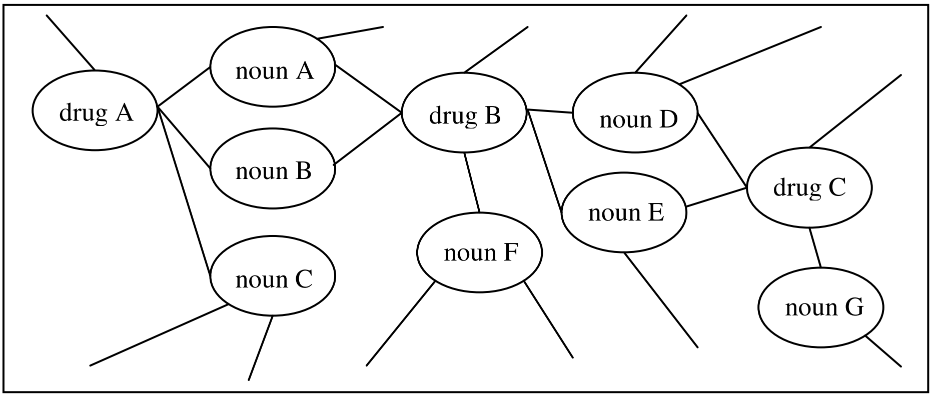


Fig. 2. An example of a network with drugs and nouns

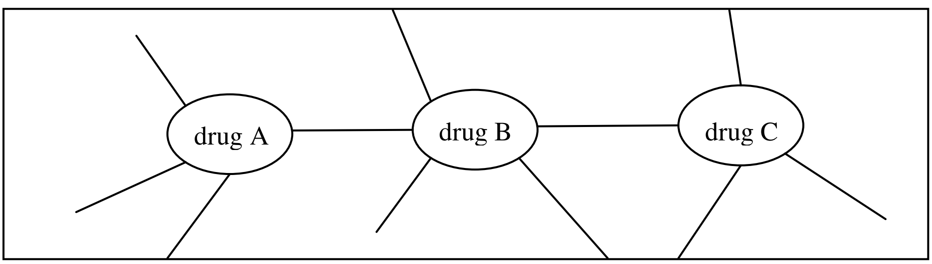


Fig. 3. An example of a network of drugs

5 Analysis 1

5.1 Method

We apply the non-hierarchical clustering method by the modularity [5] to the adjacency matrix and extract clusters connected as the same efficacy. The modularity is one of the indexes for detecting the communities from the network. If a connection between nodes in a community in the network is dense and a connection between the communities is sparse, the modularity takes the high value. This method using modularity classifies the drugs based on the network of the drugs.

5.2 Results

As a result of non-hierarchical clustering, the drugs were divided into three clusters. The numbers of drugs that belong to each cluster were 104, 72 and 57. Table 1 shows the top five frequencies of nouns that appear in the drugs included in each cluster.

Table 1. Frequencies of appearance of nouns in the top five cases in each cluster

	Cluster 1		Cluster 2		Cluster 3	
	Noun	Frequency	Noun	Frequency	Noun	Frequency
1	脳 (brain)	99	血症 (blood symptom)	45	循環 (circulation)	48
2	梗塞 (infarction)	91	血圧 (blood pressure)	43	閉塞性 (obstructive)	48
3	障害 (failure)	70	症 (symptom)	28	動脈 (artery)	48
4	頭部 (head)	55	透析 (dialysis)	28	硬化症 (sclerosis)	48
5	外傷 (injury)	55	腎不全 (kidney failure)	28	障害 (failure)	42

5.3 Discussion

Since many nouns about the head, such as “brain” “head” and so forth, appeared in Cluster 1, which is the largest cluster, we expected that Cluster 1 would be independent as circulatory organs drugs related to the head.

However, because the number of drugs in Cluster 2 is 72, and the frequency of the nouns that appear most in the cluster is 45, it is desirable to subdivide Cluster 2 including the small clusters that we should subdivide in Cluster 2. Therefore, since we cannot subdivide Cluster 2 by using the non-hierarchical clustering method, we apply the hierarchical clustering method for analysis based on inclusion relations between the clusters.

6 Analysis 2

6.1 Method

We apply the hierarchical clustering method for the connection matrix with the remaining nouns and the product names made in Section 4.2. We use the Euclid distance for distances between the drugs. Then, in order to demand a cluster expecting that it has drugs having similar efficacy, we calculate the entropy using the following expressions after the cluster division. The entropy is an index to express disorder of the information.

$$S = -\sum_{i=1}^{M}\sum_{W} P_{con}(w) \times \left(P_W^i \log P_W^i + (1 - P_W^i)\log(1 - P_W^i)\right) \tag{1}$$

P_W^i is a ratio of the elements including noun "w" in cluster i[th] after the division, and $P_{con}(w)$ is a ratio of the number of noun "w" for that of all nouns. "M" is the number of clusters. If the number of the clusters is two after the division, we calculate the entropy as M=2. If the number of the clusters is 1 before the division, we calculate the entropy as M=1. Then we calculate the information gain. The larger the information gain, the better the division. Therefore, we repeatedly divide the clusters until the information gain decreases in comparison with the division before 1.

6.2 Results

We show the results in Fig. 4 and Table 2. Figure 4 is a dendrogram showing the information gains of each division. Table 2 shows the number of elements and therapeutic category names of drugs in each cluster. As a result of the hierarchical clustering, the drugs were divided into five clusters. The numbers of drugs belonging to each cluster were 154, 32, 32, 8 and 8.

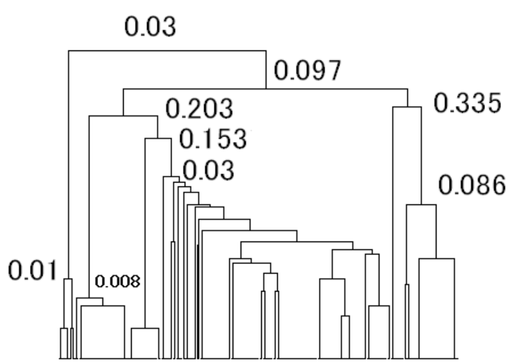


Fig. 4. Results of the hierarchical clustering method. The numbers in the figure show information gains.

Table 2. The number of elements and therapeutic category names of drugs in each cluster

Cluster number	Number of elements	Therapeutic category names of drugs in each cluster
1	154	<include many drugs with different efficacy>
2	32	"disturbance of consciousness / pancreatitis treatment drug" "Brain metabolism improvement drug" etc.
3	32	"prostaglandin E1 drug" etc.
4	9	"antithrombin drug" etc.
5	8	"circulatory disease improvement drug" "antihypertensive drug" etc.

6.3 Discussion

Since drugs having various different efficacies still coexisted in Cluster 1 including 154 elements, we could not extract therapeutic category names that plainly express the efficacy of drugs in the cluster. On the other hand, since the other clusters have therapeutic category names having almost the same meaning in each cluster, we could extract the efficacy of drugs in each cluster. According to a result such as "circulatory disease improvement drug" based on concrete efficacy or "prostaglandin E1 drug" based on the ingredient name of the drug, therapeutic category names are decided from plural viewpoints, as a therapeutic category itself. Therefore, it is necessary to unify these viewpoints of therapeutic category.

7 Conclusion

In this study, we analyzed the classification system focusing on "drugs for other cardiovascular disease" to contribute to the improvement of therapeutic categories, which have not been revised for 20 years. In order to classify drugs having similar efficacy, we proposed a method of clustering the connection of the product names of drugs and nouns included in the statements of "effect-efficacy." When we applied a non-hierarchical clustering method, there was a cluster that should be subdivided because the cluster includes small clusters. However, since we could not subdivide such cluster by the non-hierarchical clustering method, we applied a hierarchical clustering method that performs analysis based on inclusion relations between the clusters.

As a result, we classified drugs into clusters including drugs of similar efficacy. However, therapeutic category names are decided from plural viewpoints, as a therapeutic category itself. Therefore, it is necessary to unify these viewpoints of therapeutic category. In future, it is necessary to analyze the other categories by applying our method and propose a new classification system.

References

1. Ministry of Internal Affairs and Communications (2011), `http://www.stat.go.jp/index/seido/syouhin/gaiyou.htm`
2. Mochidsuki, M.: Revision; How to read package inserts to understand drugs definitely, Jiho (2004)
3. Pharmaceuticals and Medical Devices Agency (2011), `http://www.info.pmda.go.jp/`
4. Notification of the Director of the Ministry of Health and Welfare Pharmaceutical Practice: The meanings of statements in ethical drug package inserts (1997)
5. Reichardt, J., Bornhold, S.: Statistical mechanics of community detection. Physical Review E 74,016110, 1–14 (2006)
6. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69, 026113 (2004)
7. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Physical Review E 69, 066133 (2004)
8. Aaron Clauset, M.E.J.: Newman and Christopher Moore: Finding community structure in very large networks. Physical Review E 70, 066111 (2004)

Clinical Communication: Human-Computer and Human-Human Interactions

Saif Khairat[1] and Yang Gong[1,2]

[1] MU Informatics Institute,
[2] Health Management and Informatics,
University of Missouri, 65201 Columbia, USA
saifkhairat@gmail.com, gongyang@missouri.edu

Abstract. In 2006, health care in the United States encompassed 14 million professionals [1]. Such a large and complex network of employees introduces many challenges and limitations to the communication process. This research is a continuation of our efforts towards further understanding clinical communication by studying previously utilized methodologies in this field and then, proposing our new approach and its innovation. Moreover, in this paper, we discuss our results from a pilot study conducted at the Pulmonary ICU at the University of Missouri Hospital. The ultimate goal of this work is to develop a practical clinical communication model that identifies influencing communication factors. This model will serve as the foundation to our long term goal of building an ontology-driven educational tool that will be used to educate clinicians about miscommunication issues and as a means to improve it.

Keywords: Human- computer interaction, Intensive care unit, Clinical communication model.

1 Introduction

The consequences and remedies of inadequate communication within health care is a core interest in the medical informatics field. Numerous attempts have examined the communication patterns between providers and patients; this work focuses on the communication patterns within the clinical team. In health care, communication is considered the backbone of many crucial tasks such as education, information dissemination, knowledge exchange, and decision making. Clinical communication plays a pivotal role in the information flow cycle, and scientific evidence shows that communication errors can cause significant morbidity and mortality rates [2]. For those reasons, strong clinical communication is paramount for better health care outcomes. In this section, we introduce our definition of clinical communication, as well as the relationship between clinical communication and patient outcomes, and finally, recommendations from literature on how to improve communication.

In this research we define communication as the exchange of ideas, messages or knowledge between two or more entities through verbal, non-verbal, written, and visual forms where entities can be individuals or technological components. Since health care includes communication through clinicians and computers and in an

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 182–191, 2011.

attempt to encompass human-computer and human-human interactions, our definition uses the term "entities" to refer to communication among both humans and technology aided-devices. Clinical communication can be categorized into the four categories mentioned in or definition.

1.1 The Impact of Communication on Health Care Quality

The quality of care is reflected through patient outcomes and health care costs. Patient outcomes are closely tied with the occurrence of medical errors and sentinel events. As the Institute of Medicine (IOM) [3] stated that inefficient communication is a significant factor in the occurrence of medical errors. Moreover, clinical miscommunication, the failure of message exchange, is reported as the main cause for 75% of medical errors and 65% of sentinel events [4-5]. With medical errors and patient safety reported as the 8th leading cause of death among Americans [3], it is clear that significant improvements to the cause of these problems, namely clinical communication, are essential.

Economically, clinical communication is directly responsible for high health care costs and unfavorable quality. The IOM reported that preventable health care-related injuries cost the economy from $17 to $29 billion annually, of which half are health care costs [4]. This remarkably high volume of injuries has a significant impact on individual's well-being as well as the overall population health status, and the efforts of researchers should be directed towards eliminating major health care challenges.

1.2 Calls for Better Communication

Over a decade has passed since the IOM report stated that the major cause behind the increasing incidents of medical error cases is attributed to clinical miscommunication [4]. However, years later, the National Healthcare Quality Report stated that the current health care status is far from the 50% reduction in medical errors that the IOM had aimed at previously [6]. Statistics reflect the impact of clinical communications on healthcare quality and patient outcomes, and these reports suggest that an improved communication structure is needed in order to reach the IOM goals.

Furthermore, individual research efforts have highlighted the significance of communication in the performance of physicians and nurses. A study aimed at understanding the main causes behind medical error instances showed that approximately 50% of all adverse events were caused by communication errors [7]. In another retrospective study, communication errors were the lead cause of mortality, twice as frequent as clinical malpractice [8]. Beyond statistics, higher communication quality in health care is inevitable in order to increase the accuracy of information exchange.

2 Methods

The field of human factors engineering, or ergonomics, is a multidisciplinary area whose goal is to improve the relationship between humans and systems [9]. Human-Computer Interaction (HCI) is considered the science of design; it targets further understanding and supporting the interaction of users with and through technology [10]. As technology advances, more challenges are generated and higher user expectations need to be met. In this section, we provide an overview of current HCI research methods, healthcare application, and means of introducing HCI methods to improve clinical communication. This research utilizes grounded theory as a systematic qualitative research methodology, which will emphasize the generation of theory from collected data. Based upon grounded theory and human factors, our plan is to focuses on generating theories from the data collected which will improve the understanding of clinical communication ultimately.

2.1 Review of Current HCI Methods

In a field with such diversity, the utilization of research methods are numerous and for that reason this research will highlight the most common research methods used in previous HCI efforts. Among the most commonly used research methods are surveys, field study, controlled experiment, and instrument development. Table 1 summarizes results from a study conducted on the intellectual development of HCI research in Management Information systems (MIS), the study focused on understanding what types of methodologies are most popular [11]. From 378 HCI papers, the research suggests that the three most utilized methods were controlled lab experiments, surveys, and field studies, followed by field experiment, instrument development, and others. Moreover, results show that the overwhelming majority of research papers used empirical research methods, in other words methods based on observation and experience. Even though many researches used only one methodology in their work, there are an increasing number of publications that utilized two methods in their work.

Table 1. Breakdown of commonly used methodologies among 378 publications

Research Method	**Number of Papers (total = 337)**	**% from total number of papers**
Controlled lab experiment	134	35.6%
Surveys	96	25.5%
field studies	47	12.5%
Study type	**Number of papers (total = 378)**	**% from total number of papers**
Empirical	342	90.5%
Non-Empirical	36	9.5%
Number of methods per study	**Number of papers (total = 337)**	**% from total number of papers**
One method	298	88.4%
Two methods	37	10.9%

Moreover, in HCI, research methods can be categorized into three categories: (1) Natural settings, (2) Artificial setting, (3) Environment independent setting [12]. The natural setting category includes methods such as case studies, field studies, and action research. These methods can generate rich data based on the first hand data as well as, the opportunity to collect first hand data increases the reliability of findings.

On the contrary, these methods can be time consuming, data collection can be challenging due to scarcity of human resources, and the generalizability of the study maybe questionable. Methods in this category can be used to evaluate new practices, test new theories, or develop hypothesis. Artificial settings tend to include laboratory experiments as its methods, the benefits of using such methodology is the ability to control the variables and also, to replicate trials. Weakness of this category can be shown in the absence of realism and in certain cases the level of generalizability is unknown. Finally, the environment independent setting category includes research methods such as surveys, applied and basic research, and normative writings. These methods are best used to collect data from large samples, product development, and building theories and frameworks. On the contrary, there is a challenge to manipulate variables, and in some cases opinions might influence outcomes which affect the truthiness of research. This research will incorporate the significant findings from previous researches to maximize the yield of our work and hence, reach the ultimate goal of this research which is to improve clinical communication.

HCI in Healthcare. In 2008, the health care industry provided 14 million job opportunities and potentially 3.2 million new jobs by 2018, which makes the health care sector the largest and most diversified industry [13]. Also, health care is one of the few sectors that directly impact the wellbeing of citizens and better health care directly results in a shift in the population health status. The health care structure integrates two core components to provide services and care: human resources and technology. New Information Technologies (IT) serve as means towards advancement however, there are calls for better technology integration. IT in health care can be introduced in two main areas: treatment process and medical records. From that perspective, HCI research methods can be applied in those domains to provide more efficient systems and effective work flow.

In 1999, IOM released a report that documented the effect of the current health system design on medical errors, namely wrong-site surgeries and transplant errors [14]. The field of Human Factors (HF) proposes new methodologies to improve the current state of health systems and hence, introduces new interventions to prevent future tragedies. Moreover, a study reports that from 182 patients who died in 12 different hospitals, 14% were reported to be preventable and had resulted in inadequate diagnosis or treatment [15]. Therefore, there is a necessity to study and improve the current system limitations.

HCI in Clinical Communication. Clinical communication is a multifactorial, complex structure that requires the delivery of information accurately, the communication process in health care uses various mediums for communication. Professionals utilize verbal, non-verbal, and technology-aided devices to communicate amongst each other. Human-Human and Human-Computer interactions are present within this process which has resulted in inconsistent accuracy levels to communication based on the utilized communication medium. Table 2 shows the mostly used communication channels in each interaction. It is evident that within each communication channel there are many factors that need to be addressed in order to enhance overall communication levels.

Table 2. Types of clinical communication channels based on type of interaction

Interactions	Communication mediums
Human-Human interactions	Verbal
	Written
	Body language
Human-Computer interactions	Information storage
	Information retrieval
	Clinical alerts
	System messages /errors

In recent years, a new generation of computer systems was introduced to health care; those systems are designed to perform more sophisticated tasks beyond basic information entry and retrieval. Those systems aimed at assisting and improving clinical decisions. Those computer-based systems, such as Clinical Decision Support Systems (CDSS) and Electronic Medical Records (EMR), have facilitated evidence-based and patient care by reducing serious medication errors [16] and enhancing the delivery of preventive care services [17]. However, about 34% of computer-based systems have shown insignificant progress in clinical practice [17]. One of the major reasons for this inefficiency is, as the use of Health Information Systems (HIS) and Computer Information Systems (CIS) increase, new medical errors are introduced. The types of errors produced by both systems differ for each type; HIS mainly keeps track of administrative issues and CIS concentrates on patient-related data such EMRs. However, many errors from both systems can be related to miscommunication. Therefore, the communication model, proposed in later sections, addresses communication limitations in both HIS and CIS.

2.2 The Proposed Communication Model

Based on systematic literature review and preliminary data collection trials, this research has proposed the first clinical communication model which highlights communication factors, processes, and communicating parties [18]. Being the first model to delineate clinical communication, we expect this model to undergo further development, modifications, and verifications as we capture data and analyze model. A major aim of our efforts is to build a comprehensive data repository that resembles most, if not all, communication forms and factors. Scarcity of data forms a challenge to this research; therefore, we have proposed an innovative data collection methodology that will increase the scope of our efforts to include Human-Human Interactions (HHI) and Human-Computer Interactions (HCI) [19].

Since the proposal of the original communication model, the research team has been successful at capturing first hand data by observing at the Pulmonary Intensive Care Unit (PICU) at the University of Missouri. The analysis of literature findings and first hand data has added significant insights to our communication model which in return added new factors to the model. Figure 1 shows the clinical communication model with new HHI and HCI factors under each category. Also, the model shows a new formulation of the exchanges of message such that during a conversation, any given message can experience a level of distortion when received at the recipient. The alteration of a message can occur in two ways, either by missing or adding details from the original message. Reasons for incomplete message transfer can be due to external noise factors, such as side conversations, or by technology-aided devices, such as pagers and computer alerts. As for the inclusion of new information or details to the original message, reasons can include the differences in tacit knowledge such as culture, training, clinical experience, and education. Among other contributing factors, the behavioral skill level of professionals impact the way of individual's communication in health care for example, we have observed that misunderstandings can occur during multi-tasking or various interruptions. Also, during HHI, the extensive use of memory is a major contributing factor to the quality of communication; the immense amounts of knowledge and patient information can be potential risks to communication breakdowns.

On the other side, the communication mechanism between users and computer systems has many underlying factors that can lead to better communication quality. With regards to communication, HCI can be divided into two main components, users and computer systems. Users utilize the system to store or retrieve information, to perform these two functionalities, users require certain computer interaction skills. Among those skills, users need to be familiar with the system in order to enter the correct data into the correct fields. Also, users must be able to correctly retrieve and interpret information; this can be acquired with high practice and extensive system training. On the contrary, the need for user-centered systems design is paramount. The science behind usability patterns and behaviors should be the underlying science behind designing and developing clinical systems. For instance, the immense utilization of system alerts in some incidents cause users to ignore those warnings and to find workarounds. Therefore, further understanding of user skills and behaviors while using any clinical computer system will directly improve clinical communication from a HCI perspective.

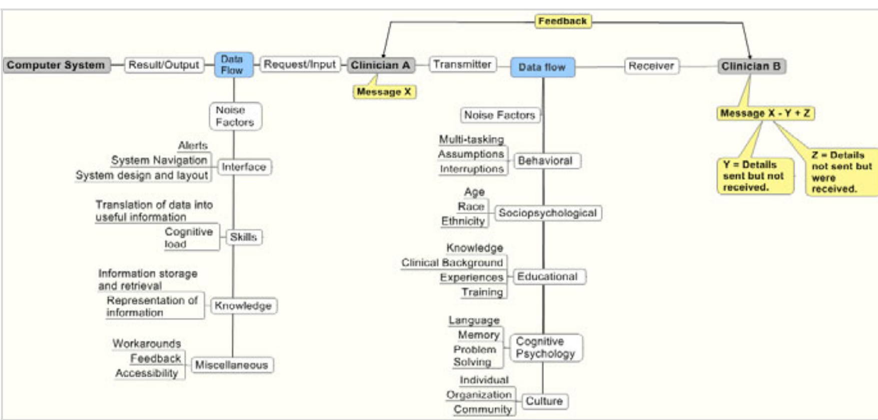


Fig. 1. Modified clinical communication model with new subcategories and message exchange representation

2.3 Approach and Environment

Since this research aims at improving communication between clinical team members, we understand that there is a need for further study and analysis of the communication behaviors of each clinical team member. For that reason, this ongoing project aims at allocating the time and efforts to observe and shadow each role within the team. By capturing role based communication information, this research will be capable of developing a clinical communication framework that will be used for education purposes for clinicians. As an initial step, we have developed a research protocol that includes of an observational study at the PICU with strong emphasis on the Lead Physician (LP). The LP is on top of the hierarchical structure and the nucleus of all communication events among the clinical team, for that reason, this research will focus on observing and studying their communication behaviors and patterns.

The observational study aims at capturing first hand data through shadowing the LP and by providing surveys to the clinical team members. The observational study will take place during the morning bed rounds in which the clinical team conducts patient visits, discusses new patient updates, and set daily goals. Through bed rounds, the researchers will utilize a checklist developed by the team to capture certain communication behaviors. After the bed rounds, the researchers will conduct a short survey with the clinical team, except for the LP, the survey questions are designed to serve as a baseline measurement for the study. Since to our knowledge there is no clinical communication checklist readily available for researchers to use, we have developed our own tool that we plan, after testing and validation, to share with the public to encourage further research.

This study will be carried out at the ICU because critical units receive patients who suffer life threatening conditions; this shows that the quality of care provided must meet the level of expectations in such critical situations. Also, there is an urgent need to improve patient safety through better understanding of communication behaviors within the ICU [20].For those reasons, we chose the ICU as the primary location for its complex and multidisciplinary nature which will allow the capturing of rare communication incidents and hence, this will increase the research generalizability. Moreover, The University of Missouri Hospital offers primary, secondary, and tertiary health care services, and it provides formal academic teaching and collaborative research services. Considered as a nationwide leader in health care quality improvement and quality patient care, the level of adherence to national health care standards and regulations are guaranteed and thus, the quality of trained staff and resources will provide a fair environment that can be extrapolated to other large health care institutes.

2.4 Tool and Data Validation

Prior to the beginning of the study, we integrated expert validity as a part of the content validity process. Expert validity is the process of obtaining feedback from domain experts on the developed tool by going through two content validity cycles at: (1) the instrumental level, (2) the item level [21]. The measurement tool was presented to domain expert to evaluate the instrument at whole by evaluating if the intended goals can be met by the design of the tool. Also, the domain expert reviewed each item of the instrument; items not recommended by the expert were eliminated from the tools to ensure higher validity levels. Also, experts made suggestions to add new items that can assist our research reach its goal by collecting the correct set of data.

Moreover, during the trial study already conducted we assessed and modified the measurement tool to best fit the hectic nature of the ICU. Through collecting first hand data, the quality of captured data will in turn increase the level of validity and reliability of our measurement tool [22]. It is reported that the best approach to increase validity is to use more than one observer, for that reason we plan to utilize two observers to conduct this study [23]. The observers will participate in the same clinical activities, use the same tools, and they will conduct the observations at the same times. Finally, we developed a set of survey questions that aim at getting a better understanding of communication behavior from the clinical team members.

3 Results

Results from the trial study suggest that there are many HHI and HCI factors that impact communication at the ICU. The factors that seem to have a major effect on the communication process included feedback, interruptions, and tasks. The quality of feedback by the LP and the frequency of interruptions appear to have significant influence on the outcomes either positively or negatively. Also, we learned that through understanding the tasks performed by the head of the clinical team, namely the LP, we can reach better understanding of the HHI. Also, we have modified the tool to capture the role of technology with respect to communication; during the trial study we observed that technology-aided devices introduced numerous amounts of interruptions during communication. For that reason, we will focus on the types and frequency of technology interruptions such as computer alerts, pagers, and telephones.

Our preliminary work in clinical communication continues along with our early studies in medical errors and communication [18, 24]. Through analyzing the strengths and weaknesses of existing methods of clinical communication, we have proposed our innovative method for examining clinical communication with an emphasis on HHI and HCI perspective, which holds promise in revealing the complexity and interruption of many technologies used in clinical communication. More efforts are needed in understanding team member interactions and will be contingent on what will be learned in the proposed study as described above.

Data collected through the procedure described above will be coded and classified into patterns and concepts with reference to the communication taxonomy proposed by Gong et al. [24]; then, concepts will be grouped into higher level categories based on the communication factors proposed in the communication model. From data analysis, we can form new communication theories based on inductive reasoning.

4 Discussion

Moving forward, the next step for this work is to start official data collection by studying LP communication habit through utilizing observational methods and by conducting survey questions. Moreover, in the future, we believe by studying various clinical team members, this comprehensive research will detect various communication patterns thus; in the future we plan to carry out more observational studies that will focus on other clinical team members including residents, respiratory therapists, or nurses. This will help refine results and uncover more factors.

One of the challenges to this research is the scarcity of data. The absence of a mandatory reporting system has resulted in many medical errors not being reported. With the exception of Veterans Health Administration and the Department of Defense, there are no nationwide reporting systems that mandate error reporting. We believe that with more data, this research can provide new information for clinicians to improve communication within a single team. Through knowledge representation, we aim at developing a framework of the communication process and hence, increase clinician's awareness of solutions and risks during interactions. Along with increasing communication awareness, these steps will enhance clinical communication, minimize medical errors, and reduce costs which will increase patient safety as well as overall population health status.

References

1. Health Care. U.S. Bureau of Labor Statistics (2008), `http://www.bls.gov/oco/cg/cgs035.htm`
2. Coiera, E.: When communication is better than computation. J. Am. Med. Inform. Assoc. 7, 277–286 (2000)
3. Kohn, L., Corrigan, J.M., Donaldson, M.: To err is human. Building a safer health system. National Academy Press, Washington, DC (2000)
4. Institute of Medicine: To Err Is Human: Building a Safer Health System. National Academy Press, Washington, DC (1999)
5. The Joint Commission Sentinel Event Statistics, `http://www.jointcommission.org/SentinelEvents/Statistics` (unavailable; last accessed July 11, 2007)
6. Agency for Healthcare Research and Quality, Rockville, MD. National Healthcare Quality Report (2006), `http://www.ahrq.gov/qual/nhqr06/nhqr06report.pdf` (accessed on May 26, 2010)
7. Bhasale, A.L., Miller, G.C., Reid, S.E., et al.: Analysing potential harm in Australian general practice: an incident-monitoring study. Med. J. Aust. 169, 73–76 (1998)
8. Wilson, R.M., Runciman, W.B., Gibberd, R.W., Harrison, B.T., Newby, L., Hamilton, J.D.: The Quality in Australian Health Care Study. Med. J. Aust. 163(9), 458–471 (1995)
9. Kantowitz, B.H., Sorkin, R.: Human Factors: Understanding People-System Relationships. John Wiley and Sons, New York (1983)
10. Carroll, J.: Human-Computer Interaction; Psychology as a Science of Design. Annual Review of Psychology 48, 61–83 (2007)
11. Zhang, P., Li, N.: The intellectual development of Human-Computer Interaction research: A critical assessment of the MIS literature (1990-2002). Journal of Association for Information Systems, 227–292 (2005)
12. Wynekoop, J.L., Conger, S.: A Review of Computer Aided Software Engineering Research Methods. In: Proceedings of the IFIP TC8 WG 8.2 Working Conference on The Information Systems Research Arena of The 1990's, Copenhagen, Denmark (1990)
13. Health Care. U.S. Bureau of Labor Statistics (2008), `http://www.bls.gov/oco/cg/cgs035.htm`
14. Kohn, L.T., Corrigan, J.M., Donaldson, M.S.: To Err Is Human: Building a Safer Health System. National Academy Press (April 2000)
15. Dubois, R.W., Brook, R.: Preventable deaths: who, how often, and why? Ann. Intern. Med. 109, 582–589 (1988)

16. Bates, D.W., Teich, J.M., Lee, J., Seger, D., Kuperman, G.J., Ma'Luf, N., Boyle, D., Leape, L.: The impact of computerized physician order entry on medication error prevention. J. Am. Med. Inform. Assoc. 6, 313–321 (1999)
17. Hunt, D.L., Haynes, R.B., Hanna, S.E., Smith, K.: Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 280, 1339–1346 (1998)
18. Khairat, S., Gong, Y.: Understanding effective clinical communication in medical errors. In: Proc. 13th Medinfo, International Congress on Medical Informatics, Stud Health Technol Inform., pp. 704–708 (2010)
19. Khairat, S., Gong, Y.: Knowledge representation in ICU communication. In: Proceedings of the 1st ACM International Health Informatics Symposium (IHI 2010), pp. 718–723. ACM, New York (2010)
20. Leonard, M., Graham, S., Bonacum, D.: The human factor: the critical importance of effective teamwork and communication in providing safe care. Qual. Saf. Healthcare 13, 85–90 (2004)
21. Hyrkäs, K., Appelqvist-Schmidlechner, K., Oksa, L.: Validating an instrument for clinical supervision using an expert panel. Int. J. Nursing Studies 40, 619–625 (2003)
22. Albrecht, T.L., Penner, L.A., Cline, R.J., Eggly, S.S., Ruckdeschel, J.C.: Studying the Process of Clinical Communication: Issues of Context, Concepts, and Research Directions. Journal of Health Communication 14(1), 47–56 (2009)
23. Berry, J.A.: Pilot study: Nurse practitioner communication and the use of recommended clinical preventive services. Journal of the American Academy of Nurse Practitioners 18, 277–283 (2006)
24. Gong, Y., et al.: Clinical communication ontology for medical errors. Studies in Health Technology & Informatics 129(Pt 2), 1007–1011 (2007)

Using Pen-Based Computing in Technology for Health

Hyungsin Kim[1], Young Suk Cho[1], and Ellen Yi-Luen Do[1,2]

[1] GVU Center & College of Computing
[2] College of Architecture
Georgia Institute of Technology, Atlanta, Georgia 30332, USA
{Hyungsin,ycho47,ellendo}@gatech.edu

Abstract. Advanced technologies open more possibilities to interact with computers in various different ways. Especially, pen-based computing provides people with an intuitive way to use a computer. In this paper, we present our research on developing pen-based neuropsychological assessment tools for older adults with cognitive impairment. First, we explain the background information and motivation to design technology for the aging population. Then, we describe our two applications: ClockReader and TrailMaker. We then discuss technological affordances to support medical assessment tasks and conclude that pen-based computing could contribute to increase the efficacy of a computer-based assessment tool.

Keywords: Pen-based Computing, Computerized Cognitive Assessments, Senior-Friendly Design, Health-Related Technology, Sketch Recognition, Human-Centered Design, Usability Clock Drawing Test, Trail Making Test.

1 Introduction

Recent advances in pattern recognition technologies and human-computer interaction push more possibilities of designing applications in pen-based computing [1]. Tablet PCs, a representative platform in pen-based computing, provide users with a natural way to interact with a computer [2]. Similar to using a pencil to write on paper, one can use a stylus to draw on top of the tablet screen. Pen-based interaction is easily accessible to people with little or no experience with computers [3]. They can leverage their previous writing experience without having to learn how to manipulate the keyboard and mouse. Despite this potential, there are only few applications utilize the unique features of pen-based interaction [1].

To leverage the opportunities of pen-based input for medical applications, we have investigated developing computerized neuropsychological assessment tools. The current practice of neurological examinations is still administered the same way as it was decades ago. Patients are asked to perform a series of assessment tasks such as diagram drawing, memory testing, or puzzle solving activities by using a pencil on a given sheet of paper. Medical practitioners such as neuropsychologists or neurologists would then spend hours analyzing and scoring the tests. The process is long and tedious. Furthermore, different administrators of the test may have different scoring criteria. By making a computer-based test, we can reduce the tedious efforts of human scoring and facilitate a consistent scoring practice and analysis [4].

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 192–201, 2011.

Furthermore, the system would provide a closer doctor-patient relationship, connecting the two more easily through telemedicine. The overarching purposes of our approach are (1) to investigate how pen-based computing technologies can play a critical role in enhancing our knowledge and understanding of the brain-behavior relationship caused by aging-related cognitive impairment; (2) to empower medical practitioners to make evidence-based decision-making; and ultimately (3) to enable transformative research in the field of neurological assessment.

In this paper, we first provide background knowledge, an overview of technology use for seniors, one of the most common diseases in the aging population, and the current use of pen-based computing in healthcare. Then, we describe the purposes of the two application developments: ClockReader and TrailMaker. Lastly, we conclude by discussing the potential of psychometric analysis with respect to pen-based computing in neuropsychological assessment.

2 Related Work

2.1 Gerontechnology

Technology has influenced the quality of our everyday lives. People have connected to one another through the Internet, mobile phones, and virtual reality systems. At the same time, the senior population has been dramatically increasing. However, studies on the design, development, and use of technology have been mostly limited to young adults, particularly aged 20-40 years old [5]. We see a need for guidelines and studies that consider the aging population when designing technology. Physical and cognitive decline of the elders require different methodologies and design considerations.

With the strong need for research on aging and technology, gerontechnology has been established as an interdisciplinary field [6]. Gerontechnology combines gerontology and technology. This approach comes from a deep understanding of the underlying characteristics of aging human beings in their social context to develop technological innovations [6]. How can technology improve the quality of everyday lives for older adults? Fisk et al. have investigated designing technology based on cognitive aging principles [5]. They argue that cognitive aging is a critical issue to be addressed. Because aging can influence task performance in several areas, technology designers should be truly aware of older adults' abilities.

Much research in gerontechnology involves health-related applications. Aiding memory, monitoring health conditions, and supporting communications with distant family members are exemplary applications to support the quality of life for older adults [6]. Indeed, health is one of the most demanding issues for older adults. In order to prevent the progression of cognitive dysfunction, older adults are encouraged to take several screening tests. In the process of developing a computerized screening tool, senior-friendly computer interaction principles should be considered in the initial stage of technology development.

2.2 Alzheimer's Disease and Related Disorders (ADRD)

A prominent public health challenge caused by aging is cognitive dysfunction, which is poor mental functioning associated with confusion, forgetfulness, and difficulty in concentrating. Alzheimer's disease is one of the representative disorders. Unlike other diseases that are physically visible, early detection of cognitive dysfunction is rarely easy. In fact, fewer than 50% of Alzheimer's cases are diagnosed, and only approximately 25% are treated, even after several years of progressive cognitive decline [7]. Unfortunately, there are no known treatments for curing Alzheimer's disease [8]. Therefore, in order to properly treat cognitive dysfunction, it is critical to identify the early process of cognitive impairment.

In 2003, the Alzheimer's Foundation of America established the third Tuesday of November as National Memory Screening Day [9]. The goal of this initiation is to promote the early detection of Alzheimer's disease and related disorders (ADRD) and to encourage timely intervention and treatment of them. It is often difficult to detect cognitive impairment because (1) it is hard to differentiate cognitive impairment from normal cognitive degeneration due to aging; (2) there is limited opportunity for seniors to meet with specialists, such as neurologists or neuropsychologists, unless they have serious observable symptoms; and (3) the disease is usually progressively developed, so it is hard to capture the appropriate moment, which normally requires continuous monitoring through everyday activities.

Therefore, our society needs a novel, quick-and-easy screening system for cognitive impairment, as well as preventing the onset of aging cognition. The significant amount of research currently being conducted regarding phone- or computer-based dementia screening indicates that dementia screening should no longer be limited to clinicians' offices [7].

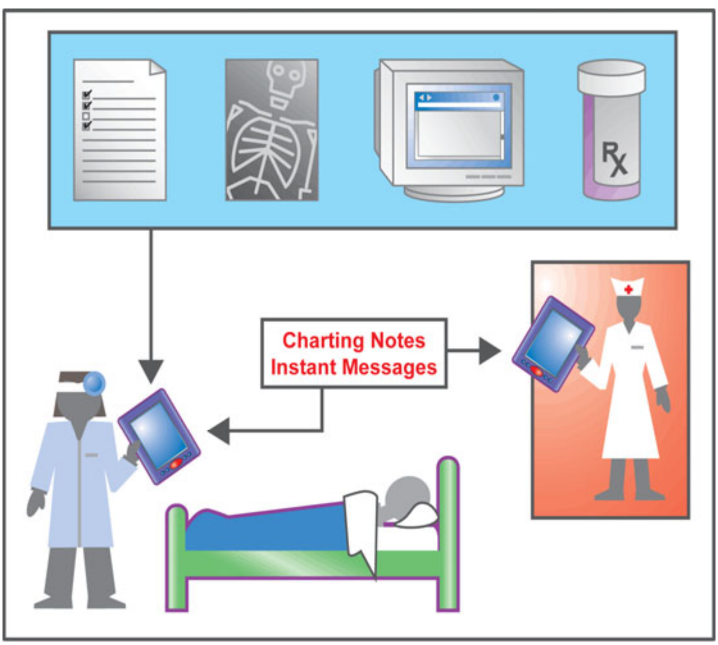


Fig. 1. Diagram showing a doctor and nurse stay connected with the Tablet PC (adopted by Microsoft White Paper 2004 [10])

2.3 The Use of the Tablet PC in Health Care

According to the white paper on the *Business Case for Tablet PC* (2004), Microsoft emphasizes that when implementing the electronic medical record system, the Tablet PC would provide great value to healthcare workers with flexible access to patient information at the point of care in the hospital, and outside of the office, such as home care [10]. They pointed out that the electronic data in the Tablet PC can help reduce medical errors, increase doctor and nurse efficiency, and shorten the duration of patient visits. Figure 1 below shows a visualization of how the Tablet PC's unique form factor and the capability to write directly on the screen would provide more effective ways to facilitate communication in patient care. We also see the potential of the use of tablets in increasing the effectiveness of administrative workflow in complex hospital environments.

3 Applications in Pen-Based Computing

The purpose of our proposed applications in pen-based computing is to develop a quick-and-easy neuropsychological assessment tool for people with age-related cognitive impairment. Based on the literature review on identification of neuropsychological issues, we decided to develop the Clock Drawing Test (CDT) and the Trail Making Test (TMT) into computerized systems.

The applications are developed in C# programming language using "Microsoft Windows XP Tablet PC Edition Software Development Kit 1.7" and "Microsoft Visual Studio 2010." For the first release, the running environment of the program is limited to the Microsoft Windows platform, equivalent to or better than "Windows 2000 Service Pack 4" with "Microsoft.Net Framework 3.5 Service Pack 1." Every coordinate of the cusps and intersections of each stroke (even if it represents a character) are stored in the memory. In this section, we introduce two applications we have developed: the ClockReader and the TrailMaker Systems.

3.1 The Clock Drawing Test (CDT)

Prior to explaining the ClockReader System, what it is and how it is used, let's first explain the Clock Drawing Test (CDT). The CDT has been used for decades as a neuropsychological screening test [11, 12]. The CDT is usually part of the 7-Minute Screen, CAMCOG (Cambridge Cognitive Examination), and Spatial-Quantitative Battery in the Boston Diagnostic Aphasia Examination [13]. The CDT focuses on visual-spatial, constructional, and higher-order cognitive abilities, including executive aspects [12]. The CDT accesses human cognitive domains from comprehension, planning, visual memory, visual-spatial ability, motor programming and execution, abstraction, concentration, and response inhibition [11, 14, 15]. The major value for clinicians to conduct this test is that the CDT can capture cognitive dysfunction and provide concrete visual references of the patients.

The Clock Drawing Test (CDT) is one of the simplest, but most commonly used screening tools to detect cognitive impairment in seniors [11]. By simply asking people to draw a clock, it easily identifies people with dementia [13]. Figure 2 shows three different clock drawings from three patients [11]. The drawings clearly show degradation of the patient's cognition. A salient pattern is that patient could not use the space of the clock evenly. Clock drawings from people with cognitive impairment frequently show missing or extra numbers, or misplaced clock hands [11, 14].

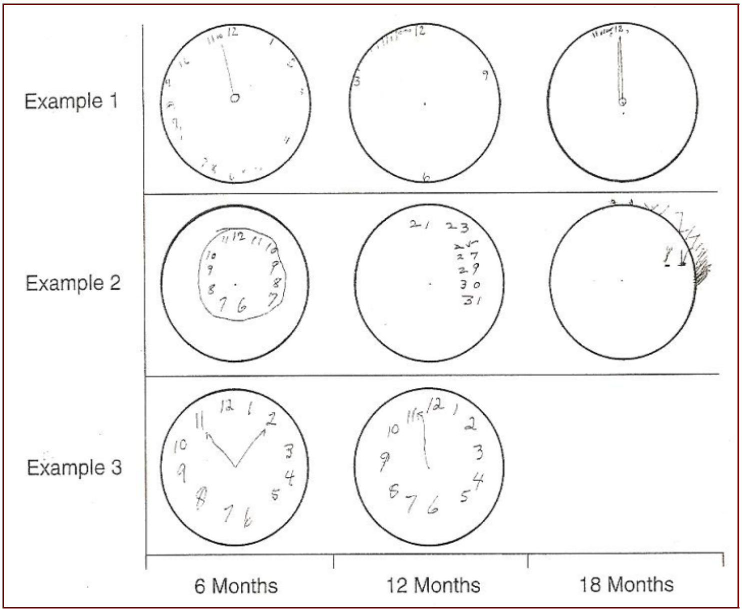


Fig. 2. Three examples of clock drawings showing deterioration in dementia [11]

There are two variations of the test. One test asks the patients to draw a clock in a pre-drawn circle. This test focuses on the spatial distribution of the numbers, as well as the hands of the clock. The other test does not provide any pre-drawn circle. Patients are asked to draw a freehand circle on the paper. In some cases, patients are shown a picture of a clock drawing and are asked to copy that onto the paper. Patients are then asked to set different times for the clock, such as 11:10, 1:45 or 3:00 [11-13].

Numerous scoring systems are also available. Each of the scoring systems places differing emphases on visual-spatial, executive, quantitative, and qualitative issues [15]. Qualitative errors can provide more valuable information to understand different patterns of drawings due to the progression of dementia. For example, clocks drawn by patients with right frontal lesions show difficulty with number position. Clocks drawn by patients with left frontal damage show reversal of the minute and hour hand proportion [11]. Overall, the CDT is accepted as the ideal cognitive screening test, based on widespread clinical use [13]. Among published studies, the CDT achieves a mean sensitivity of 85% and a specificity of 85% [16, 17].

3.2 The ClockReader System

The purpose of the ClockReader System is to enable patients to take the Clock Drawing Test without the presence of a human evaluator. The system consists of three main components: data collection, sketch recognition, and data analysis. First, the system would record and recognize a patient's freehand drawing and collect the data. Then, based on the scoring criteria, the system would automatically analyze the drawing and report the score.

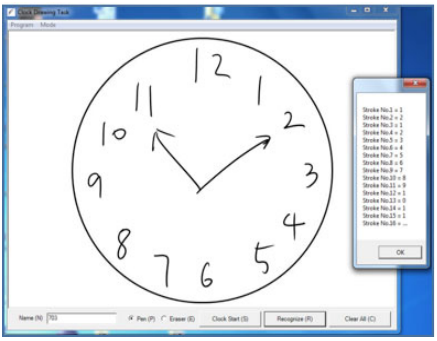


Fig. 3. A Screen shot of the ClockReader System

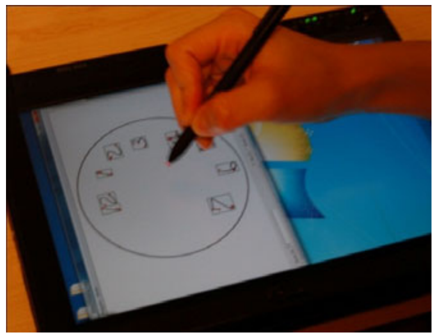


Fig. 4. A picture showing how a patient uses the ClockReader System

Figure 3 shows a screen shot of our ClockReader system. The small window on the right shows the results of the digit recognition. Figure 4 shows a patient drawing a clock in the ClockReader system using a stylus on a tablet PC.

Let us briefly explain how the ClockReader recognizes handwriting. The ClockReader system is developed based on the Microsoft Tablet PC recognizer. The current Tablet PC SDK provides character recognition through the stroke level. However, rather than capturing each character based on a stroke, we modified it to capture the character level. Some Arabic numbers include more than two strokes For example, the number 4 is often written with two strokes. Therefore, the recognition process of the ClockReader System starts from setting a rectangular area per character, passing the data from the rectangular area to the Microsoft SDK handwriting recognition engine, and then finally saving the recognized results as a string. We also implemented a simple machine learning technique, a Context-Bounded Refinement Filter Algorithm, to improve the recognizer accuracy [4]. After the recognition process, the program then analyzes the relative position between each number and scores the complete clock drawing with the given criteria.

3.3 The Trail Making Test (TMT)

The Trail Making Test (TMT) is also frequently used to assess psychomotor speed, complex attention, and executive functions for people with cognitive impairment [12]. Historically part of the Army Individual Test of General Ability, the Trail Making

Test assesses general intelligence [18]. There are two different versions of the TMT: Part A and Part B. Both versions ask a subject to connect the dots of 25 consecutive targets on a sheet of paper. In Part A, the TMT includes only numbers as a target (such as 1, 2, to 25). Part B of the TMT includes both numbers and letters as targets (such as from 1, A; 2, B; to 13, L). Due to the complexity of the test, Part B of the TMT is more frequently used to access prefrontal dysfunction, which is observable by performance in flexible shift response sets [19]. The goal of the TMT is to measure how quickly the subject completes the test without errors. Usually, people with aphasia or detectable neglect show latency in completing TMT Part B [18].

3.4 The TrailMaker System

The TrailMaker System enables patients to take the Trail Making "Part B" Test without the presence of a human evaluator. Compared to the ClockReader System, the TrailMaker System includes a very simple scoring criterion, such as whether the dots are connected with the appropriate sequences (such as 1-A-2-B-3-C). First, when a user draws a line to connect each big circle dot, the evaluation algorithm initiates to detect the correctness. Then, based on the amount of correctness, the system would automatically analyze the connecting-the-dots result and report the score. The system provides the user with an erase-all function to delete all the lines they drew and restart the test. Thus, users can restart as many times as they like. However, the number of trials will be recorded together with the total scoring results and the completion time. Figure 5 below shows a screen shot of our TrailMaker system.

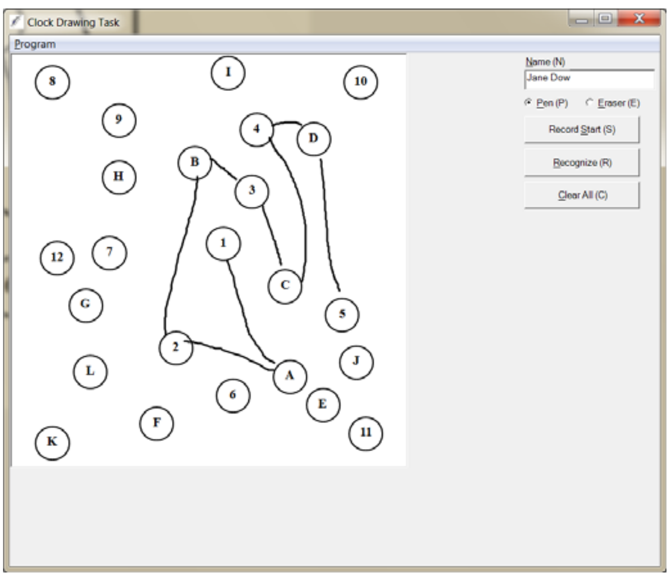


Fig. 5. A Screen Shot of the TrailMaker System

4 Technological Affordances

Researchers advocate that pen-based computing can be the next generation of new interaction techniques for seniors [1]. A recent study shows that older adults have better interaction with a touch screen and digital pen compared to indirect manipulation devices, such as the mouse and keyboard [3]. Unlike WIMP (Windows, Icons, Menus, and Pointers)-based interaction, a pen can provide the seniors with more direct interaction. Users can put the stylus on the tablet screen and see the cursor directly below it. This is contrast to using a relative pointing device such as a mouse, which requires hand-eye coordination (moving the mouse on the desktop while looking at the screen to find the cursor location).

To provide the affordance of a paper-and-pen environment for patients, we implemented the system as a Tablet PC application. Our preliminary usability test results show that using a stylus on the surface of a Tablet PC is similar in form to using a pen on a piece of paper [20]. Many users commented that it is actually easier to draw on the Tablet PC than on a piece of paper. We found that even a 91-year-old grandfather who had never used a computer before succeeded in completing the Clock Drawing Test using our system without any difficulties. However, we believe that the opportunities to leverage pen input can go beyond simple recognition and its physical affordance.

The research focus should shift from what and how to recognize to how to interpret the recognized data for meaningful use. The pen-input enables us to measure more diverse visual spatial factors of human handwriting. More importantly, the computerized screening system will make it possible to gather behavioral data, such as airtime, tendency to pause, patterns of exerting pressure, and sequence of the drawing. Thus, based on our work, we propose applications in pen-based computing to perform in-depth psychometric analysis in terms of **Processing Capturing** and **Airtime Capturing**.

Instead of analyzing the drawing as a final output, Kaplan argues that the process-oriented approach can be the best way to understand a patient's performance, and she later names it as the Boston Process Approach, quantifying the process as a natural evolutionary step [15]. **Processing Capturing** can be useful in interpreting qualitative information rather than simply reporting result scores. For example, the ClockReader provides information to capture a patient's drawing process and records the drawing sequence; therefore, a clinician can play back and interpret the planning strategy of a patient (e.g., 12, 3, 6, 9, or, 1, 2, 3, 4). This provides clinicians with new and useful information that was not previously available from paper-and-pencil tests. Furthermore, it can also be used to make a differential diagnosis on the sub-types of Alzheimer's disease and related disorders by categorizing the different error types.

Both systems collect two types of time capture records: (1) total completion time and (2) airtime. The total completion type is always a good standard measurement in any kind of testing. The **Airtime Capturing** can provide important data for clinicians to understand a patient's cognition. For example, the TrailMaker captures "airtime," which is the time when the patient is not drawing – the time of pausing. When patients connect the dots, at a certain point, they may hesitate to connect, perhaps due to memory problems. None of the existing criteria for the paper-based Trail Making Test take this factor into consideration. However, airtime could be a useful indicator of abnormal and unstable cognition. The airtime graph from a patient can provide time-related patterns of the patient's drawing task. That information can show critical moments when the patient spent time thinking before performing executive behaviors.

5 Conclusion and Future Directions

As the senior population increases, more dementia screening and prevention support will be needed for patients and medical practitioners. In this paper, we discussed two pen-based computing applications to support neuropsychological assessment, especially for seniors with cognitive impairment. For future work, we will first add several modules to enhance the systems. Then, we plan to deploy our systems in community centers or clinics to for longitudinal study. Ultimately, we expect to contribute to the efficacy of a computer-based cognitive screening test by leveraging the potential of pen-based computing. We believe that technology can offer more effective and efficient cognitive impairment screening. Furthermore, we see the potential of extending this research in building neuropsychological assessment systems to identify the early detection of children with developmental disorders.

Acknowledgement. This research would not have been possible without the support of all the senior volunteers and researchers at Emory Alzheimer's Disease Research Center. We appreciate their time and insightful feedback for our study.

References

1. Grosky, W.I., Zeleznik, R., Miller, T., van Dam, A., Li, C., Tenneson, D., Maloney, C., LaViola, J.J.: Applications and Issues in Pen-Centric Computing. IEEE MultiMedia 15, 14–21 (2008)
2. Davis, R.: Magic Paper: Sketch-Understanding Research. Computer 40(9), 34–41 (2007)
3. Arias-Torres, D.: The Design and Evaluation of a Pen-Based Computer Interface for Novice Older Users. In: Seventh Mexican International Conference on Computer Science, ENC 2006, pp. 142–150 (2006)
4. Kim, H., Cho, Y.S., Do, E.Y.: Context-Bounded Refinement Filter Algorithm: Improving Recognizer Accuracy of Handwriting in Clock Drawing Test. In: Visual Representations and Reasoning Workshop of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010) W12, Atlanta, GA, USA, pp. 53–60 (2010)
5. Charness, N., Park, D., Sabel, B.: Communication, Technology and Aging: Opportunities and Challenges for the Future. Springer Publishing Company, Heidelberg (2001)
6. Burdick, D., Kwon, S.: Gerotechnology: Research and Practice in Technology and Aging. Springer Publishing Company, Heidelberg (2004)
7. Solomon, P.R., Murphy, C.A.: Should we screen for Alzheimer's disease? A review of the evidence for and against screening Alzheimer's disease in primary care practice. Geriatrics 60, 26–31 (2005)
8. Desai, A.K.: Healthy Brain Aging: Evidence Based Methods to Preserve Brain Function and Prevent Dementia, An issue of Clinics in Geriatric Medicine. Saunders (2010)
9. Powers, R.E., Ashford, J.W., Peschin, S.: Memory Matters. Alzheimer's Foundation of America, AFA (2008)
10. Microsoft: Business Case for Tablet PC (2004), http://www.microsoft.com/downloads/en/details.aspx?FamilyID=4F47B606-BD1D-4638-853D-3C1A9736915B&displaylang=en

11. Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K., Delis, D.C.: Clock Drawing: A Neuropsychological Analysis. Oxford University Press, USA (1994)
12. Lezak, M.D., Howieson, D.B., Loring, D.W.: Neuropsychological Assessment. Oxford University Press, USA (2004)
13. Strauss, E., Sherman, E.M.S., Spreen, O.: A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press, USA (2006)
14. Ismail, Z., Rajji, T.K., Shulman, K.I.: Brief cognitive screening instruments: an update. Int. J. Geriatr. Psychiatry 25, 111–120 (2010)
15. Kaplan, E.: The process approach to neuropsychological assessment of psychiatric patients. Journal of Neuropsychiatry 2, 72–97 (1990)
16. Shulman, K.I.: Clock-drawing: is it the ideal cognitive screening test? Int. J. Geriatr. Psychiatry 15, 548–561 (2000)
17. Shulman, K.I., Feinstein, A.: Quick Cognitive Screening for Clinicians: Mini Mental, Clock-drawing and Other Brief Tests. Informa Healthcare (2003)
18. Strauss, E., Sherman, E.M.S., Spreen, O.: A compendium of neuropsychological tests: administration, norms, and commentary. Oxford University Press, US (2006)
19. Goldstein, G., Incagnoli, T.M.: Contemporary Approaches to Neuropsychological Assessment. Springer, Heidelberg (1997)
20. Kim, H., Cho, Y.S., Guha, A., Do, E.Y.: ClockReader: Investigating Senior Computer Interaction through Pen-based Computing. In: CHI Workshop on Senior-Friendly Technologies: Interaction Design for the Elderly, Atlanta, GA, USA, pp. 4513–4516 (2010)

Using a Smart Phone for Information Rendering in Computer-Aided Surgery

Gaël Le Bellego[1], Marek Bucki[1,2], Ivan Bricault[1], and Jocelyne Troccaz[1,*]

[1] UJF- Grenoble 1 / CNRS / TIMC-IMAG UMR 5525,
Grenoble, F-38041, France
jocelyne.troccaz@imag.fr.
[2] CHU Grenoble, Radiology and imaging Department,
Grenoble, F-38043, France
first-name.last-name@imag.fr

Abstract. Computer-aided surgery intensively uses the concept of navigation: after having collected CT data from a patient and transferred them to the operating room coordinate system, the surgical instrument (a puncture needle for instance) is localized and its position is visualized with respect to the patient organs which are not directly visible. This approach is very similar to the GPS paradigm. Traditionally, three orthogonal slices in the patient data are presented on a distant screen. Sometimes a 3D representation is also added. In this study we evaluated the potential of adding a smart phone as a man-machine interaction device. Different experiments involving operators puncturing a phantom are reported in this paper.

Keywords: distant display, smart phone, physical interface.

1 Introduction

1.1 Computer-Aided Surgery (CAS) Principles

For more than two decades, navigation systems are proposed to the clinicians to assist them during an intervention [1]. Typically CT anatomical data are collected for a given patient before the intervention. These data allow planning the intervention, for instance by defining a target position for a surgical instrument or for a prosthesis element. During the intervention, the navigation system gives information to the clinician about the progress of the intervention: typically the position of the instrument relatively to the target and to pre-recorded anatomical data is visualized in real-time. The position of surgical instruments in space is known thanks to a tracking device called "localizer." Most often, an additional stage is required to bring the surgical plan recorded pre-operatively to the intra-operative conditions; this stage is named registration. The approach is very similar to navigation assistance of cars. The localizer is similar to the GPS positioning system and the CT data correspond to the recorded road and city maps on which the position of the car is displayed.

* Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 202–209, 2011.

Such navigation systems are also called “passive” assistance systems [2] since they only render information to the clinician who can use it in the way he/she wants. Alternative “active” assistance systems exist: in this case a robot can perform autonomously a part of the intervention. Intermediate “semi-active” solutions also exist where a robot may be tele-operated by the clinician or a programmable mechanical guide may constrain the possible motion of the instrument moved by the clinician. This paper focuses on passive navigation systems.

1.2 Man-Machine Interaction in CAS

One very conventional way of displaying guidance information to the clinician is based on the dynamic visualization of orthogonal slices computed in the volume of recorded data (see figure 1). A sagittal slice is a vertical slice which passes from front to rear dividing the body into right and left sections. The transverse slice (also called the horizontal slice or axial slice) is obtained by cutting the volume by a plane that divides the body into superior and inferior parts. It is perpendicular to sagittal slices. A coronal slice (also named frontal slice) is a vertical slice that divides the body into ventral and dorsal section. The intersection point of the three slices generally corresponds to the tip of the navigated instrument. When the instrument is moved the three slices are updated accordingly.

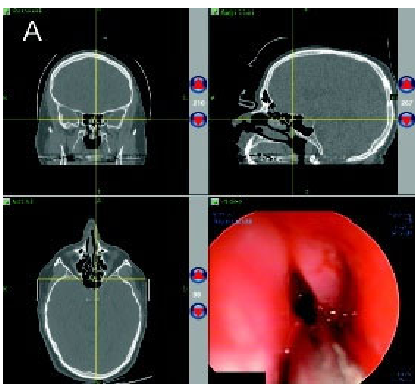


Fig. 1. Typical display of a navigation system. Coronal slice (top left), sagittal slice (top right), transverse slice (bottom left), endoscopic view. The position of the surgical instrument tip is visualized by the intersecting yellow crosses on each of the three slices. (Source: Neurosurgery Focus © 2008 American Association of Neurological Surgeons).

Because a meaningful representation of the tool trajectory is generally very important, the standard cutting planes presented above can be replaced by what is called pseudo-slices. A pseudo-transverse slice includes the tool axis and is slightly angulated with respect to a conventional transverse slice (see figure 2 left). Figure 2 right shows the GUI (Graphical User Interface) of a navigation system for punctures. A pseudo-transverse slice and a pseudo-sagittal slice help visualize the position and orientation of the puncture needle with respect to the patient anatomy.

When a target trajectory has been predefined in a planning stage, some additional information may be presented to the user, in order to compare the executed trajectory with the planned one. The trajectories can be visualized using specific visor displays in addition to the slice viewer. [3],[4] and [5] propose such “targeting” interfaces in their GUI.

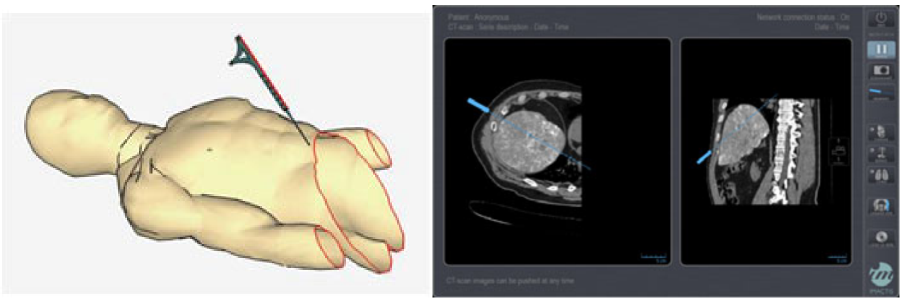


Fig. 2. Pseudo-slices. Definition of the pseudo-transverse slice (left). Two slices (pseudo-transverse and pseudo-sagittal) in a computer-assisted needle puncture software (right).

Most often data are displayed on a 2D screen installed in the viewing space of the clinician. Using the displayed information generally requires moving the surgical instrument without looking at it or at the patient. Perceptual continuity [6] is no longer guaranteed. Augmented reality systems have been proposed to remove this limitation. They are based on semi-transparent devices such as Head Mounted Displays [7] where navigation data is overlaid on intra-operative images given by an already existing sensor (for instance a surgical microscope [8]). Except for this last case, very few augmented reality systems are used in routine clinical practice.

More recently several groups [9], [10] proposed to display part of the guidance information on small mobile screens where the displayed data may depend on the position of the screen. [11] and [5] proposed to attach the screen to the instrument. [5] presents some experimental evaluation of different display modes.

A few years ago, thanks to the technology evolution in particular regarding PDAs and smart phones, our team decided to explore this potentially new interaction paradigm for CAS applications. The purpose of this work was to study the feasibility of using a mini-screen, within a close range to the operating site, in order to display partly or totally the guidance information to the clinician during interventions. Different combinations of displays and different representations of data and interaction modes with the data were explored for interventions such as punctures. The experimental environment and the conducted experiments are presented and discussed in the following sections.

2 Material and Methods

2.1 Experimental Environment

The system (cf. figure 3) includes the standard elements of a navigation system: the optical localizer (passive Polaris from NDI, Inc.) enables tracking in real time objects equipped with reflecting markers. It is linked to the computer running the navigation application. A 19 inches screen is connected to the computer; in the following we will call it the master screen. As regards the mini-screen, several possibilities were envisioned (LED, OLED, LCD screens, PDA, smart phones). We selected the iPhone3G which advantages were to have many embedded features (good quality display, wi-fi communication, accelerometers, tactile interaction, camera, microphone, etc.), a complete development environment and a large interest and experience from the HCI community. A client-server application controls the dialog between the main computer and the smart phone.

The user can interact with the navigation application on the main computer and master screen in a traditional way (scrolling, mouse clicking, etc.). When using the smart phone the interaction with the data is possible using the tactile screen (scrolling for zooming functions, clicking for definition and recording of a position of interest) and the accelerometers (for navigation around a point of interest).

The experiments are performed using a custom-made phantom. A block of deformable PVC in which the punctures are performed is placed inside a manikin. The puncturing instrument and the manikin are equipped with reflecting markers and are tracked by the localizer. CT data are associated to the phantom for navigation.

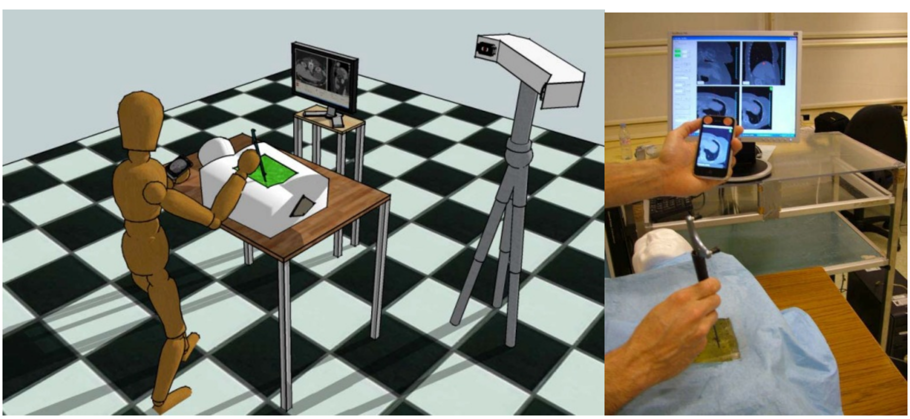


Fig. 3. Experimental set-up

2.2 Data Representation and Operating Modes

After having experienced several possible representations of data with users, we selected three of them:

- "Triple-ortho" representation: this is the standard representation used for navigation; three orthogonal slices (transverse, sagittal and coronal) intersecting at the tip of the instrument are computed in the 3D pre-operative CT data.
- "Double-pseudo" representation: two pseudo-transverse and pseudo-sagittal slices including the instrument axis are computed in the CT pre-operative 3D data. A "single-pseudo" (pseudo-transverse) counterpart is also used when displayed on the smart phone.
- "Adjustable pseudo" representation: a single pseudo-transverse slice is computed in the 3D pre-operative CT data. Although the presented information is still defined by the instrument position, its orientation can now also be freely adjusted by the operator around the tool axis or a marked position for a deeper exploration of the data close to the instrument.

In order to evaluate the ability to delegate part of the GUI to the smart phone, we compared four solutions:

- "Standard mode": a triple-ortho or double-pseudo representation is displayed on the master screen only.
- "Double mode": a triple-ortho or double-pseudo is displayed on the master screen and a pseudo-transverse image is displayed on the smart phone; the smart phone view is also added to the master screen.
- "Remote mode": an adjustable-pseudo representation is displayed on the smart phone only. The accelerometers control the orientation of the slice around the tool axis or around a marked position.
- "Distributed mode": a standard double-pseudo representation is displayed on the master screen. At any time the user can record the tool position. Then, while the master screen keeps displaying the standard double-pseudo, the user can navigate through the data around the recorded position using the smart phone which displays the adjustable slice.

2.3 Experiments

Experiment n°1. Three experimental conditions were tested: (TO) standard-mode with triple ortho representation of data, (TO+iP) double mode with triple-ortho on the master screen and pseudo-transverse on the smart phone, (TO+iPA) distributed mode with triple-ortho on the master screen and adjustable pseudo-transverse on the smart phone. 30 operators were involved: 12 clinicians and 18 non clinicians (PhD and Master students). Training was performed before the experiment with the set-up presented in section 2.1 and a synthetic CT dataset. After training the dataset was replaced by a real exam of a patient having a quite big and easily detectable renal cyst; the target was the cyst; the user could scan the exam before starting the punctures. The three conditions were presented in a random order to the operator. Speed of execution of the puncture and rate of success are recorded. Between two exercises, a 5mn rest was left to the user. After 10mn of unsuccessful trial, the puncture was considered as a failure. After each experiment the operator had to fill a questionnaire (about his/her fatigue, cognitive effort, liking of the tested solution with Likert scales from 1 to 7) and was eventually asked to give a ranking of the three solutions.

Experiment n°2. A second experiment was set where only pseudo-slices were used for representation. Three experimental conditions were tested: (DP) standard mode with double-pseudo representation of data, (iPA) remote mode with an adjustable pseudo-transverse on the smart phone, (DP+iPA) distributed mode with double-pseudo on the master screen and adjustable pseudo-transverse on the smart phone. The incremental nature of the representation involved that the preliminary training is performed in this specific order. But the experiment itself was here again performed with a random order. After the training the dataset is replaced by a real exam of a patient; the target is a simulated hepatic cyst placed in a delicate anatomical area; the user could still scan the exam before starting the punctures. 6 operators (all clinicians) contributed to this experiment. The order of tested conditions, time condition for failure, recorded parameters and questionnaires were similar to experiment n°1. The distance from the tip of the instrument to the target was recorded when the user considers that it had been reached.

3 Results

Comparisons of the three conditions in both experiments used a non parametric test (Friedman test) with paired samples (in each experiment, each operator experimented three conditions for the same task). For experiment n°1 where two populations were involved (clinicians and non clinicians), comparisons between the two populations used the Mann-Whitney test.

3.1 Experiment n°1

For the global population of 12+18 subjects, the felt comfort, felt cognitive effort and felt fatigue are similar in the three conditions. The duration of targeting for TO+iPA is in average longer than for TO which is longer than TO+iP; only the difference between TO+iPA and TO+iP is statistically significant. The liking of the interaction mode is in average lower for TO than for TO+iP which is lower than for TO+iPA; the liking of TO is significantly different from the other two conditions. Regarding the ranking of the interaction modes, in average TO+iP was preferred to TO+iPA which was preferred to TO; only the difference between TO+iP and TO is statistically significant.

When focusing on the clinician subgroup, all the measured or felt characteristics were similar in the three conditions. When focusing on the non clinician subgroup, the felt comfort, the felt cognitive effort, the felt fatigue and the ranking are similar in the three conditions. The measured duration is in average lower for TO than for TO+iP which is lower than for TO+iPA; the duration for TO+iPA is significantly different from the duration for the other two conditions. In average the liking of TO is lower than the liking for TO+iP which is lower than the liking for TO+iPA; the liking for TO is significantly different than the liking for the other two conditions.

When comparing the two subgroups for the measured duration and for liking of the interaction modes, no difference could be exhibited between the subgroups. This probably means that the difference between statistics for the global population as compared to statistics for the subgroups is due to the number of people involved.

3.2 Experiment n°2

Regarding the measured duration, measured accuracy, felt cognitive effort, and felt fatigue, no significant difference was computed among the three conditions. The felt comfort of the iPA condition was significantly different (lower) than the felt comfort of the other two conditions. The liking of the interaction mode was in average lower for iPA than for DP which was lower than DP+iPA; only iPA was significantly different from DP+iPA. Finally, for the ranking of the three types of interaction modes, all the subjects placed the iPA in the last rank; in average DP+iPA was the best placed, before DP and iPA. A significant difference between iPA and DP+iPA could be exhibited.

4 Discussion and Conclusion

In the first experiment, the increase of duration measured for the TO+iPA condition probably comes from the time spent to explore locally the data with the adjustable slice. Other experiments would be necessary to determine if any particular stage of the puncture is preferably concerned with this adjustment (initial orientation of the needle? fine approach to the target? other?). The fact that TO is in average longer than TO+iP could be explained by the fact that the smart phone with the pseudo-inverse slice adds some useful information that makes the puncture easier with respect to the TO representation alone. However since the difference was not statistically significant this explanation has to be taken with special care and specific additional experiments would be needed.

As concerns the interface distribution on the two displays, the users did not appreciated the remote mode (experiment n°2) where guiding information is only present on the smart phone and they felt uncomfortable with it. The size of the display and the resolution of the displayed data were mentioned by the users as the main limitations. The available zooming function was however nearly never used although systematically introduced.

The combination of displays in a distributed mode was generally appreciated in the conducted experiments. From our point view, the master screen brings good quality information enabling a global 3D perception of the conducted task; however the visualized data directly depend on the position and orientation of the tool. While this is particularly useful for the initial orientation of the needle before entering the tissues, the progress of the needle limits any further exploration of the data. This is probably where the smart phone is the most useful. It allows a local exploration around the current tool position for instance for controlling the absence of anatomical obstacles or the presence of remarkable anatomical features for an easier and safer access to the target.

One limitation of this work is the relatively small number of subjects. The fact that the experiments were rather long due to training, multiple conditions, associated interviews and filling of questionnaires was an obstacle to the recruitment of clinicians even though we moved to the hospital to make their involvement as short and as easy as possible. Having more clinicians would allow classifying them in terms of their expertise of surgical navigation. It would also permit to draw more definitive conclusions about the best representation and interaction modes. Other applications with different type of assistance could also certainly benefit from such integrated technology.

However we think that the presented experiments show that the use of a mini-screen for CAS guidance is feasible, well accepted and is probably a good complement to a larger screen.

References

1. Reinhardt, H.F.: Neuronavigation: a ten year review. In: Taylor, R.H., Mosges, R., Lavallée, S. (eds.) Computer-Integrated Surgery, pp. 329–341. MIT Press, Cambridge (1996)
2. Cinquin, P., et al.: Computer Assisted Medical Interventions at TIMC Laboratory: passive and semi-active aids. IEEE Engineering in Medicine and Biology magazine, special issue Robots in Surgery 14(3), 254–263 (1995)
3. DiGioia, A.M., Simon, D.A., Jaramaz, B., Blackwell, M., Morgan, F., O'Toole, R.V., Colgan, B., Kischell, E.: HipNav: Pre-operative Planning and Intra-operative Navigational Guidance for Acetabular Implant Placement in Total Hip Replacement Surgery. In: Proceedings of CAOS 1995, Bern, Switzerland (1995)
4. Chavanon, O., Barbe, C., Troccaz, J., Carrat, L., Ribuot, C., Blin, D.: Accurate guidance for percutaneous access to a specific target in soft tissues. Journal of Laparoendoscopic and Advanced Surgical Techniques 9(3), 259–266 (1999)
5. Kassil, K., Stewart, A.J.: Evaluation of a tool-mounted guidance display for Computer-Assisted Surgery. In: Proceedings of CHI 2009, Boston, pp. 1275–1278 (2009)
6. Dubois, E., Nigay, L., Troccaz, J.: Assessing perceptual continuity and compatibility in augmented reality systems. International Journal on Universal Access in the Information Society, special issue on Continuous Interaction in Future Computing Systems 1(4), 263–273 (2002)
7. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. ACM SIGGRAPH Computer Graphics 26(2), 203–210 (1992)
8. Edwards, P.J., King, A.P., Hawkes, D.J., Fleig, O., Maurer Jr, C.R., Hill, D.L., Fenlon, M.R., de Cunha, D.A., Gaston, R.P., Chandra, S., Mannss, J., Strong, A.J., Gleeson, M.J., Cox, T.C.: Stereo augmented reality in the surgical microscope. Stud. Health Technol. Inform. 62, 102–108 (1999)
9. Weber, S., Klein, M., Hein, A., Krüger, T., Lueth, T.C., Bier, J.: The navigated image viewer – evaluation in maxillofacial surgery. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 762–769. Springer, Heidelberg (2003)
10. Mansoux, B., Nigay, L., Troccaz, J.: Abstract and concrete interaction with mixed reality systems. The case of the mini-screen, a new interaction devices in computer-assisted surgery. In: International Workshop on Exploring the Design and Engineering of Mixed Reality Systems, IUI-CADUI 2004, Funchal, Portugal (2004)
11. Lavallée, S., Troccaz, J.: Localized medical instrument with tilt and swivel screen. PCT Patent WO2004/080323, French patent 03/029402 by Université Joseph Fourier and Praxim

A Proposal of Contraindication Database for Medicines

Ryo Okuya[1], Hirotsugu Ishida[2], Keita Nabeta [2], Masaomi Kimura [1],
Michiko Ohkura [1], and Fumito Tsuchiya [3]

[1] Shibaura Institute of Technology,
3-7-5 Toyosu, Koto Ward, Tokyo 135-8548 Japan
[2] Graduate School of Technology, Shibaura Institute of Technology,
3-7-5 Toyosu, Koto Ward, Tokyo 135-8548 Japan
[3] International University of Health and Welfare,
2800-1 Kitakanemaru, Ohdawara City, Tochigi 324-8501 Japan
{l07035, masaomi,ohkura}@shibaura-it.ac.jp,
{m110013,m709102}@shibaura-it.ac.jp, ftsuchiya@iuhw.ac.jp

Abstract. In recent years, despite various measures taken to reduce medical accidents as a result of confusions over drugs, cases of medical malpractice have occurred in Japan. As a countermeasure supported by a Health Labor Sciences Research Grant in 2009, drug information databases based on drug package inserts have been created for computer systems to prevent accidents caused by incorrect treatment of drug information [1]. However, the data in the databases remains problematic. In this study, we propose data item sets to be defined in drug information databases.

Keywords: Medical safety, Package inserts, Databases.

1 Introduction

Safe usage of medicines is one of the keys to preventing medicine-related accidents. One attempt to increase safety is a device to provide medicine-related data for doctors to verify whether the selection of medicines is appropriate for the purpose of treatment. In Japan, computerized prescriber order entry (CPOE) systems are widely used to prescribe medicines, and there have been many efforts to prevent the wrong input of medicine names: improved order in the list of medicines, highlighting frequently confused medicine names by adding certain symbol characters and so forth. In addition to such efforts, it is important to provide suitable medicine-related information for doctors in order to recognize the selection errors of medicines. With this in mind, we have conducted studies to create databases containing the information required for a CPOE prescription checking system.

In 2009, such databases for Japanese medicines were proposed with the support of a Health Labor Sciences Research Grant. The databases consist of tables such as the one storing contraindication data. The contraindication data is one of the key data described in package inserts to prevent the occurrence of fatal medical accidents. In spite of their effort to define table schema and their importance, the data therein are less structured and difficult to utilize in CPOE. This is because they inherit

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 210–218, 2011.

characteristics of original package inserts, namely, the diversity of structures and expressions of descriptions. Original package insert data, though their contents are officially defined by the authorities, contain information expressed as sentences, tables, figures without strict constraints of expressions. Since the package inserts are an official source of medicine information, the contraindication part of the database is made based on them and are described in sentences or phrases. For such less constructed data, we need to employ a full-text search technique to find whether contraindication data include some keywords. Imagine the case that some medicine contraindicates use in ‘infants two months or younger’. If you confirm that the medicine has a contraindication to infants, you have to compare the string ‘infant’ with each word in the contraindication phrase. In order to search an exact match of keywords effectively, we have to turn the data into fully-structured ones based on code systems. The current version of the databases does not satisfy this requirement.

In this study, we decompose the phrases and the sentences in the contraindication part of the database into fundamental sets of keywords and group them to design code systems. For some tables, we decompose them by hand. Additional to these analyses, we also investigate the correspondence of terms defined in some other master data, e.g. ICD-10, the master data of diseases [2]. We separate the terms in the master data from the contraindication data and define the new attributes corresponding to the remaining parts. Based on the results of these analyses, we propose a database schema that is necessary for storing (structured) contraindication data.

2 Target Data

Our target databases are those developed with the support of a Health Labor Sciences Research Grant as outlined below.

- Contraindicated patient-type database
  - The database of contraindication descriptions related to patients such as elderly people, pregnant women, nursing women and babies (**Table 1**). The data are the summaries of precautionary statements and contraindication statements in package inserts that are related to such patients and are described in a few phrases. The databases are problematic, since they lump many kinds of contraindication information together. This occurs because various aspects of contraindication remain to be unarranged. For example, there exists the data ‘babies suffering from exomphalos’ and ‘babies two years or younger’. The former is the condition related to disease and the latter is related to age. Additional to this, plural conditions can co-exist in one record representing one medicine. In order to realize a search for an exact match, the data first need to be turned into normal form, namely, be broken into separate records for each condition. Several different spellings of terms are also problematic.
- Contraindicated disease database
  - The database of contraindication descriptions related to patients suffering from a certain disease (**Table 2**). The data are the summaries of contraindication statements in package inserts that focus on diseases except for their reasons and

are described in a few phrases. The problem in this database is that not only disease/symptom names but also other information, such as administration duration of other medicines (e.g. phrases like ‘during administration of catecholamine’) are contained in it, though it is intended to be a list of disease/symptom names. This occurs because the administration information is indirectly related to the disease information. As is in the contraindicated patient database, various kinds of contraindication information are also lumped together and several different spellings of terms co-exist.

Table 1. Sample data in the contraindicated patient-type database

Product names	Baby-related information	Nursing-woman-related information
Mercurochrome liquide	Babies suffering from exomphalos	
Selapina granule	Babies 2 years or younger	Nursing women (long-term continuous application)
Diapp suppository 4	Low-birth-weight baby, newborns	

Table 2. Examples of data in the contraindicated disease database

Product names	Contraindicated disease names
Prorner tablet 40	bleed
Junsi Ephedrine hydrochloride	during administration of catecholamine
Dilute iodine tincture. OY	iodine hypersensitivity
Calvital	hypersensitivity to iodine

3 Methods

3.1 Analysis of Contraindicated Patient-Type Database

First, we manually split the data stored in the database into phrases whose meaning can be interpreted. Then, we classified the phrase into a set of words representing each meaning. Phrases are split manually because words obtained by application of techniques such as morphological analysis often cannot be interpreted as medical information. **Fig. 1** shows an example of how to split the data. We classified the words with confirming the package inserts published by the Pharmaceuticals and Medical Devices Agency [3]. Since there are different expressions in the database, though they have the same meaning, we collected such expressions to create a dictionary and to standardize them.

2歳未満の幼児 (Infants under two years old)	→	2歳未満 (Under two years old)	幼児 (Infants)

Fig. 1. An example of the data split

3.2 Analysis of the Contraindicated Disease Database

Since some data stored in this database is not a standard disease, we investigated whether the data match a disease name in standard disease master data published by The Medical Information System Development Center. We classified each of the data into four types of matching. The first type is data exactly matching one of the disease names in the master data, and is called *exactly matched*. The second type is data matching a combination of two or more words that are disease names or their modifiers in the master data, called *semi-exactly matched*. **Table 3** shows samples of the data in this type. The third type is data only whose substring matches a disease name or its modifier in the master data, called *partially matched*. The fourth type is data that are not applicable to these types, called *unmatched*.

We analyzed data in each type as follows.

We confirmed whether the exactly matched data is actually a disease name.

We investigated what the additional information to the semi-exactly matched or partially matched data is other than disease name.

Since the *unmatched* data do not match any disease names in the master, we investigated what information the data contain.

Table 3. Samples of the *semi-exactly matched* data

A phrase in data	Substrings matching the master data
急性冠動脈疾患 (Acute coronary artery disease)	急性(Acute)
	冠動脈疾患(Coronary artery disease)
重症糖尿病 (Severe diabetes)	重症(Severe)
	糖尿病(Diabetes)
特発性血管腫瘍 (Idiopathic vascular tumor)	特発性(Idiopathic)
	血管(Vascular)
	腫瘍(Tumor)

4 Results

4.1 Results of the Analysis of Contraindication Patient-Type Database

By classification of the phrases in the data in this database, we found that they contain not only information about medicine administration and the condition of patients but also the words/phrases representing patient categories (**Table 4**). We also found that the information about medicine administration can be categorized as either prohibited method, duration, body parts, dosage, purpose or frequency, and that the information about medicine administration can be categorized as either prohibited method, duration, the body parts, dosage, purpose or frequency. Moreover, the information was also categorized as either reason or condition of prohibited administration. The information about patient conditions was classified into prohibited age, duration, state, and disease

Table 4. Sample of patient categories

Patient category	Words/phrases	
高齢者(Geriatric)	高齢者(Geriatric)	老齢(Old age)
授乳婦(Nursing woman)	授乳婦(Nursing woman)	授乳中(Nursing)
乳幼小児 (Babies and children)	小児(Child)	幼児(Infant)
	新生児(Neonate)	患児(Diagnosed infant)
妊婦(Pregnant women)	妊婦(Pregnant women)	産婦(Parturient women)
	妊娠中 (During pregnancy)	妊娠又は妊娠している可能性 (Pregnant or possibly pregnant)

Table 5. Samples of the information about medicine administration

Categories	Phrases	
方法(Method)	密封法 (Occlusive dressing technique)	点滴静注 (Intravenous infusion)
期間(Duration)	長期 (A long term)	長期間 (A long term)
部位(Body parts)	指(Finger)	子 宮 膣 内 (The insides of vagina of womb)
投与量(Dosage)	大量 (Large dose)	過量投与 (Overdose)
目的(Purpose)	広範なテスト(Extensive test)	子宮卵管撮影 (Hysterosalpingography)
頻度(Frequency)	頻回 (Frequent)	
理由(Reason)	ヒスタミン遊離 (Histamine release)	生殖発生毒性試験 (Reproduction test)
条件(Condition)	リバビリンとの併用 (Ribavirin combination)	

Table 6. Samples of the conditions of a patient

Categories	Phrases	
年齢(Age)	生後3カ月未満 (Under three months of age)	2歳以下 (Two years or younger)
期間(Duration)	3カ月以内 (Within three months)	12週未満 (Less than twelve weeks)
	初期 (Early phase)	末期 (Last phase)
状態(State)	胎位異常 (Fetal malpresentation)	継続が危険 (Continuation is dangerous)
病名(Disease name)	腸アトニー (Intestinal atony)	高度黄疸 (Severe jaundice)
	高ビリルビン血症 (Hyperbilirubinemia)	肺動脈閉鎖 (Pulmonary atresia)

Table 7. A List of sets of synonyms

Combination of the similar phrases	
乳幼児(Infants)	乳・ 幼児(Infants)
妊婦又は妊娠している可能性 (Pregnant or possibly pregnant) 妊婦及び妊娠している可能性 (Pregnant and possibly pregnant)	妊娠又は妊娠している可能性 (Pregnant or possibly pregnant) 妊娠又は妊娠の疑われる (Pregnant or doubt of pregnant)

name. **Table 5** and **Table 6** show samples of the information about medicine administration and the condition of the patient and their categories.**Table 7** shows a list of synonym sets. The phrases in the same cell are synonyms that have similar meanings such as '妊娠'(pregnant) and '妊婦'(pregnant woman), '可能性'(potential) and '疑われる'(in doubt). Such combinations must be unified in a word.

We propose data items that should be included in a table schema based on the classification results and combination of similar phrases. The data items are as follows:

— A patient category
— An administration method
— An administration period
— A target body part
— Dosage
— Frequency of administration
— The purpose of administration
— The reason for contraindication
— Patient age
— Patient state
— Duration of patient state
— A disease name

Additional to the definition of these data items, we created a master table for each of them, each of whose data has ID that is identical to their meanings but whose branch number indicates difference of expression (**Table 8**). Based on these data items and codes, we created a database whose samples are shown in **Table 9**. In order to avoid sparse assignment of data in each record, the column named *Contraindication condition codes* stores the data in items other than patient categories.

In order to evaluate the data items newly proposed in this study, we randomly selected 1,000out of 7,152 records in the contraindicated patient database and confirmed that the table can store the information included in them without omission.

We also found that our schema do not admit a query with multiple conditions to match, since it assumes each piece of contraindication information is dealt with independently. From the viewpoint of safety, this might be sufficient to cover necessary contraindication information, since a database user should decide whether he/she adopts its restrictive conditions.

Table 8. Sample codes related to contraindicated patient types

Codes	Contraindicated patient types
A001-1	高齢者 (geriatric)
A002-1	老齢 (old age)
B001-1	授乳婦 (nursing women)
B001-2	授乳中 (nursing)
B001-3	授乳期 (lactation)
C001-1	小児 (a child)
C002-1	新生児 (an infant)
C003-1	低出生体重児 (low-birth-baby)
C003-2	未熟児 (immature baby)

Table 9. Sample data in the resultant database

YJ code	HOT9 code	Patient cat.	Contraindication condition codes
1119401A1036	100316101	C001-1	P1005-1 (2 years or younger)
1119401A1036	100316101	C004-1	P1005-1
7219413A1023	111867401	A001-1	null

4.2 Results of Analysis of the Contraindicated Disease Database

Since all of the exactly matched data were disease names, we can deal with them as *contraindicated disease names*.

Semi-exactly matched data were found to contain not only disease names but also their additional information such as the extent of the disease, frequency, body parts, and the conditions of a patient, as is shown in **Table 10** and **Table 11**.

Table 10. Samples of information that include the extent of disease

Phrases	**Disease names**	**The extent of disease**
軽症熱傷(mild burn)	熱傷(Burn)	軽症(Mild)
高度肝障害(Severe hepatopathy)	肝障害(Hepatopathy)	高度(Severe)

Table 11. Samples of information that include the conditions of a patient

Phrases	**Disease names**	**The Conditions of a patient**
妊娠ヘルペス (Gestational herpes)	ヘルペス(Herpes)	妊娠(Pregnant)

The partially-matched data and the unmatched data also contained additional disease information such as the extent of diseases, body parts, the conditions of a patient, reactive drugs, medical devices, treatments and related diseases. Some data did not contain disease but the condition of patients, reactive drugs and treatments. We should note their categories, and they must be dealt with as distinct items in the database. Their samples are shown from **Table 12** to **Table 14**.

Table 12. Samples of reactive medicines

Phrases	Timings	Drugs reactive with the target
経口避妊薬服用中 (During taking oral contraceptive pill)	服用中 (During taking)	経口避妊薬 (Oral contraceptive pill)
MAO阻害剤投与中 (During taking MAO inhibitor)	投与中 (During taking)	MAO阻害剤 (MAO inhibitor)

Table 13. Samples of treatments

Phrases	Treatments
血液透析中(During hemodialysis)	血液透析(Hemodialysis)
PUVA療法実施中(During PUVA therapy)	PUVA療法(PUVA therapy)

Table 14. Samples of the conditions of patients

Phrases	The condition of a patient
白血球数2,000 / mm3以下 (The white blood cell count is 2,000 per cubic millimeter or less)	白血球数2,000 / mm3以下 (The white blood cell count is 2,000 per cubic millimeter or less)

We propose the data items to describe contraindicated diseases as follows:

- Drug ID
  - YJ code
  - HOT9 code
  - A product name
- Contraindication information
  - A contraindicated disease name
  - A contraindicated patient state
  - A contraindicated treatment
  - A contraindicated drug reactive with the target drug
- Additional information
  - Details of a disease
  - The extent of a disease
  - The phase of a disease
  - Reactive drugs
  - A body part
  - A treatment
  - A medical device
  - A causative disease and/or symptom

- A causative action
- A disease not eligible
- An accompanying disease

5 Conclusion

In this study, in order to contribute to designing databases for a CPOE prescription checking system, we investigated the databases for Japanese medicines storing contraindication data. The target databases were the contraindicated patient database, the contraindicated disease database and the contraindicated drug combination database.

As for the contraindicated patient-type database, we confirmed that many kinds of contraindication information related to patients are lumped together in it and that several different spellings of terms exist therein. We also found that we can decompose and sort the contraindication information in each data into items such as medicine administration information, patient-related conditional information and patient categories. The medicine administration information is categorized into an administration method, the body part to be applied and quantity. The patient-related conditional information consists of age, disease names and so forth.

Although the contraindicated disease database is expected to be a list of contraindicated disease/symptom names of each medicine, it also includes medical treatment, patient status and mutual interaction between medicines. As for contraindicated disease/symptom names, we found that there is additional information such as the parts of the body, the extent and period of disease and so forth. We defined data items to store these and also created master data for each kind of information.

Since the data in these databases are described in natural language, there is variance in the expressions of terms. In order to unify the expressions, we identified terms that have the same meaning and defined a coding scheme to reflect the identification on.

In the future, we will extend our database schema to enable processing of more flexible queries by reflecting the correspondence relationships between pieces of contraindication information. We will also implement contraindication databases other than the contraindicated patient database.

References

1. Fumito, T., Masaomi, K.: Studies on the utilization of computerized drug package inserts. Summary Research Report 2009, Health Labor Sciences Research Grant (2009)
2. The Medical Information System Development Center, `http://www.medis.or.jp/`
3. Pharmaceuticals and Medical Device Agency, `http://www.info.pmda.go.jp/`

Results of the Usability and Acceptance Evaluation of a Cardiac Rehabilitation System

Cecilia Vera-Muñoz[1], María Teresa Arredondo[1], Ignacio Peinado[1], Manuel Ottaviano[1], José Manuel Páez[2], and Arturo Díaz de Barrionuevo[3]

[1] Life Supporting Technologies
{cvera,mta,ipeinado,mottaviano}@lst.tfo.upm.es
[2] ETSI Telecomunicación
jm.paez@upm.es
[3] ETSI Montes,
Universidad Politécnica de Madrid,
Ciudad Universitaria s/n. 28040, Madrid, Spain
a.dbarrionuevo@upm.es

Abstract. Cardiac rehabilitation programs are extremely important during the recovery phase of patients who suffered a Myocardial Infarction (MI). Traditionally, these programs aimed at recovering the cardiovascular functioning by means of tailored exercise programs. However, during the last years, researchers and practitioners have started to consider a multi disciplinary approach for the interventions, where patient education and health behaviors changes play a central role both in rehabilitation and in secondary prevention. Nevertheless, the lack of resources and the need to have health professionals continuously involved in the patient supervision process, does not make possible to extend rehabilitation programs to all potential patients in the health care system. This paper presents the results of the usability and acceptance validation of a personalized home based cardiac rehabilitation system developed under the framework of HeartCycle, a project partially funded by the European Commission.

Keywords: Usability and acceptance evaluation, cardiac rehabilitation.

1 Introduction

Myocardial Infarction (MI) is a leading cause of morbidity and mortality in the world. According to the World Health Organization (WHO), an estimated 17 million people died from cardiovascular disease in 2005, 7.2 million of them due to heart attacks. If current trends are allowed to continue, it is estimated that 23.6 million people will die from cardiovascular disease by 2030, mainly from heart attacks and strokes [1]. The American Heart Association and American College of Cardiology recommend following cardiac rehabilitation programs after suffering a myocardial infarction, as they increase patients' chances of survival. These programs are specifically designed to help patients' recovery after a heart attack, from other forms of cardiovascular disease or after surgery to treat heart diseases. They are often divided into phases

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 219–225, 2011.

involving monitored exercise, support and education about lifestyle, all designed to help patients regain strength, prevent their condition from worsening and reduce their risk of future heart problems [2].

This paper presents the results of the usability and acceptance validation of a system aimed to support the cardiac rehabilitation process of coronary artery disease patients after suffering a myocardial infarction. The system has been developed within the scope of the HeartCycle project [3], partially funded by the European Commission under the 7th Framework Programme.

HeartCycle applies ICT to provide patients who suffered a heart attack with a complete solution that helps and supports them during the recovery phase. It offers a comprehensive cardiac rehabilitation program and personalized plans designed to be followed by patients at home. The solution not only includes an exercise plan for physical recuperation, but also tackles potential risk factors and lifestyle changes through innovative education and motivation techniques.

2 Technical Approach

HeartCycle solution comprises an elaborated cardiac rehabilitation program supported by a technological platform that includes systems both for patients and health care professionals. The professional system helps physicians to supervise patient's evolution during the rehabilitation process by allowing them to set up personalized care plans, regularly update them and conduct the appropriate follow up on their progress. Care plans encompass the prescribed exercise sessions for the whole rehabilitation period, including details on frequency, duration and level of effort to be achieved for each of the phases.

The patients system consists of three different devices: a PC application or patient station, a portable unit (PDA) and a sensor for monitoring vital signs. The patient station includes a complete set of functionalities that they can use to successfully accomplish the complete rehabilitation program. More specifically, patients can utilize this application as a communication channel with health professionals (by means of messages that are prompted in their PC screen), to schedule and personalize

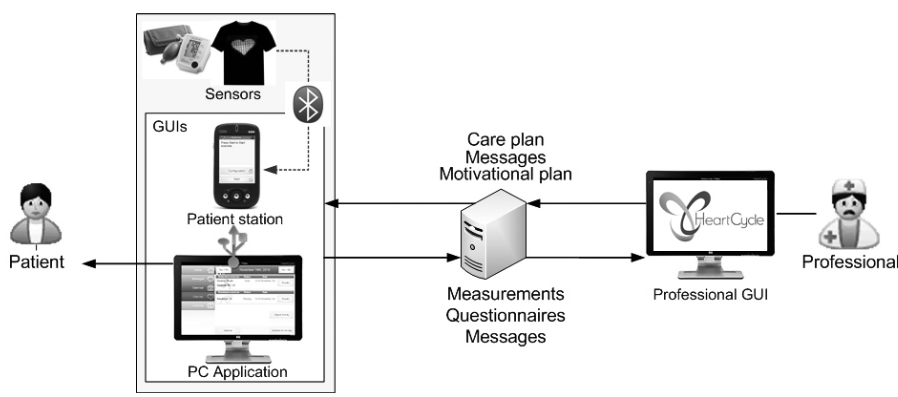


Fig. 1. Patient systems overview

their exercise sessions, and to receive feedback on the achieved progress. Additionally, the system will also motivate them towards their goal achievement, by presenting rewarding messages and providing them with personalized and appropriate educational content (Fig. 1).

During the exercise sessions, patients wear a specific sensor for collecting vital signs and a portable unit (PDA) that guides them during the different stages of the session. These devices continuously control the health status of the patients ensuring that they are always on the adequate and safe range levels while performing the prescribed exercise. Once an exercise session is finished, all the gathered information is transferred to the patient station and, afterwards, to the professional system, where the physicians can evaluate the performance and adjust the care plan, if needed.

3 Methodology

The patient system has been developed following the principles of User-Centered Design (UCD), aiming to guarantee its acceptance and usability. The methodological approach chosen has been Goal Oriented Design (G-OD) [4]. G-OD proposes an iterative approach that starts with a thorough research of the problem under study and continues with intermediate validations of increasingly complex prototypes with actual and potential users of the system (namely, with patients and health professionals during the development life cycle of the system) [5]. In the case of the patient system, the development started with the analysis of the specific use case aimed to reflect all patient's needs and goals, and the innovative features that should be provided. This study led to the identification of four main blocks of functionalities to be considered: (1) a care plan, with the prescribed exercise sessions, including details on frequency, duration and level of effort to achieve; (2) messages system, to establish a communication with the patient; (3) reminders of pending activities and feedback on the progress of the completed sessions; and (4) a complete motivation and education strategy focused on increasing the knowledge of patients about their disease and rehabilitation process, while incentivizing them to follow the exercise program.

The concept also included personalization features, a requirement of paramount importance in order to improve the user experience and to facilitate the patient's adherence to the system. Different ways of personalization have been considered in the interaction: personalization of educational content, personalization of feedback or form factor, and personalization based on knowledge or technical skills. Finally, a specific design of the graphical user interfaces for the envisaged system, including an attractive look and feel, was depicted.

Once the initial system design was finalized, a mock-up was built and validated with patients by means of interviews, using questionnaires that tackled motivation, usability and acceptability aspects. Questionnaires combined semantic differentials and Likert scales, used to gather quantitative data regarding usability and acceptability of the system, with open questions aimed to obtain qualitative data about the patients' insights with respect to motivational and acceptability issues [6]. Patients' interviews were conducted in two different countries (United Kingdom and Spain) and locations (Hull and East Yorkshire Hospitals NHS Trust Castle Hill

Hospital, and Hospital Clínico San Carlos Madrid), involving a total number of 33 patients, suffering either coronary artery disease or heart failure.

Each interview itself included three main parts with a total duration of 45 minutes. The first ten minutes were devoted to a general introduction of the purpose of the test, the signing of the informed consent, and to gather demographic data and general health status of the patients. Afterwards, patients were asked to answer several questions focused on general motivation factors, more specifically on physical exercise and rehabilitation, medication, and learning preferences. The interview then continued with the usability and acceptability parts. After a short introduction and demonstration of the application, patients were asked to use it and perform specific tasks. Finally, they filled in a questionnaire with their impressions, as well as a scoring sheet [7].

4 Results

The outcomes of the validation were quite positive. System acceptance questionnaires showed that most of the patients (81%) liked the system and considered they would use it frequently (81%). Also, the majority of them (85%) thought that the application would help them to control their health and would motivate them towards a healthier lifestyle (67%). Only 11% of interviewees believed that the system would invade their privacy, thought that it would make them feel neglected by their physician, or

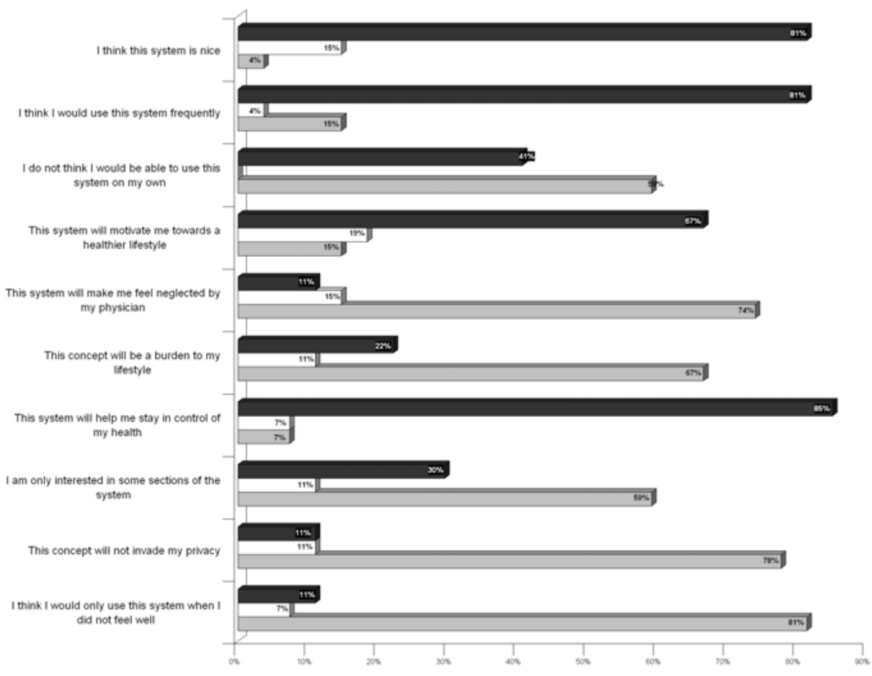


Fig. 2. Acceptance questionnaire results (27 Patients interviewed). Rating values from 1 (completely disagree) to 5 (completely agree). Results analysis disagree (1 or 2), neutral (3), agree (4 or 5)).

reported they would just use it when they would not feel well. Moreover, approximately one third of the patients (30%) would like to select the features to use of the application. On the other side, a considerable percentage of patients did not feel capable of using the system on their own (41%), the majority of them were advanced age users living in Hull (UK). In these cases, interviewers informally asked the patients about the reasons for this concern, which mainly resulted to be reluctance to use new technology because they were not familiar with it (Fig. 2).

Usability aspects of the system were also very well valued (Fig. 3). A great majority of patients perceived it as usable (76%) and stimulating (79%). The presented application was found very pleasant by 88% of the interviewees and also 88% of them rated it as not intimidating. Finally, the great majority of patients (94%) considered the system very interesting.

Although no major problems were identified in the interaction with the application, 12% of the interviewed patients rated usability aspects with low values, mainly due to their advanced age and their reluctance to use new technologies. In these cases, patients provided the interviewers with some suggestions and feedback that helped the designers to improve subsequent versions of the user interfaces (i.e. simplifying the contents of certain screens, adding supplementary pictures for better understanding the functionality, etc.)

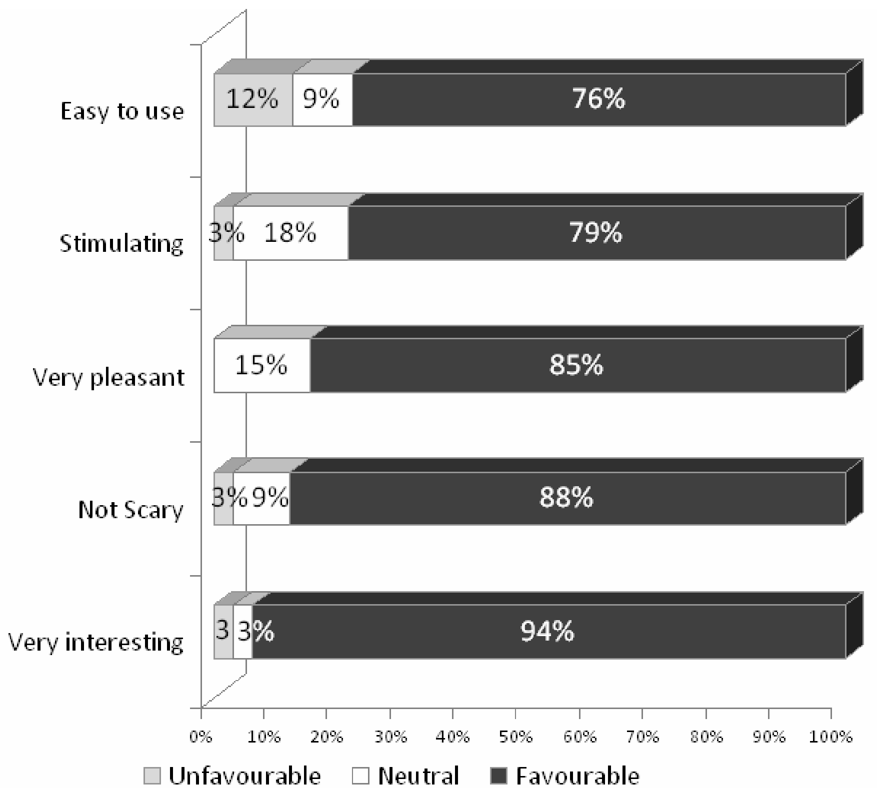


Fig. 3. Usability questionnaire results (33 patients interviewed): Rating values from 1 to 5. Results unfavorable (1 or 2), neutral (3), favorable (4 or 5).

The validation of the system also included an initial study on the possible education and motivation material to be included in the application. The results revealed differences between the patients' interests on the different topics and the need of introducing personalized education and motivation strategies and content in the system.

The outcomes of this validation led to a refinement of the global design and the implementation of an improved cardiac rehabilitations system. This new version has taken into consideration all the received suggestions and has specifically addressed the aspects that obtained worse rates in the conducted validation. Moreover, a new and more attractive look and feel has been included in the final development.

The complete system for cardiac rehabilitation is planned to be validated in a clinical trial during 2011, involving approximately 60 patients in 3 different locations (Spain, United Kingdom and Germany).

5 Conclusions

Although cardiac rehabilitation programs have proven to be really important for a complete recovery after suffering a heart attack, the lack of resources in the health care system and the need of continuous health professional supervision have usually hampered their introduction into the standard care. Nowadays, only a reduced number of post-MI patients can access this kind of programs, which are often offered in certain specialized health centers.

The presented system propose a personalized home based rehab program that constitutes an attractive alternative to the traditional ones, as well as represents an interesting solution for possibly extending cardiac rehabilitation to a higher number of potential patients. As it offers patients supervised care while they are following a program at home, it could be easily adopted and implemented within the health care system with a limited amount of extra resources.

The positive outcomes of the conducted validation tests, both in terms of usability and acceptance, have proved that the solution could be a feasible option to be incorporated into the normal health practice and have also set up very good perspectives for the adoption of this kind of systems in the future. It is expected that HeartCycle solution will highly benefit a great number of post MI patients in their recovery process.

Acknowledgments. We would like to thank the whole HEARTCYCLE Project Consortium for their valuable contributions for the realization of this work, especially to the NHS Trust Castle Hill Hospital (Hull and East Yorkshire Hospitals –United Kingdom) and Hospital Clínico San Carlos Madrid (Spain). This project is partially funded by the European Commission.

References

1. World Health Organization. Fact sheet N°317. updated (September 2009)
2. Hamm, L.F.: Core Competencies for Cardiac Rehabilitation/Secondary Prevention Professionals. Journal of Cardiopulmonary Rehabilitation and Prevention (2010) update, `http://www.jcrpjournal.com` (last access 4th February 2011)

3. Heart Cycle project, Compliance and effectiveness in HF and CHD closed loop management, European Commission, FP7 – ICT216695, http://heartcycle.eu/ (last access February 2011)
4. About Face 3.0 The Essentials of Interaction Design. Wiley Publishing, Inc., Chichester (2007); ISBN-10: 0470084111
5. MyHeart: Fighting cardiovascular diseases through prevention and early diagnosis. European Commission, FP6 – IST-507816, http://www.hitech-projects.com/euprojects/myheart/ (last access February 2011)
6. Tullis, T., Albert, B.: Measuring the user experience. Morgan Kauffmann, San Francisco (2008); ISBN: 9780123735584
7. Heart Cycle project, Deliverable 11.4 – Patient Loop Specifications and User Manuals, Annex of CAD Patients Confrontation Results (February 28, 2011)

Construction and Analysis of Database on Outer Cases of Medicines

Hironori Yoshimi[1], Hiroki Muraoka[1], Akira Izumiya[1], Masaomi Kimura[1], Michiko Ohkura[1], and Fumito Tsuchiya[2]

[1] Shibaura Institute of Technology, Japan
[2] International University of Health and Welfare, Japan
{l07144,l06112,m708101,masaomi,ohkura}@sic.shibaura-it.ac.jp, ftsuchiya@iuhw.ac.jp

Abstract. This study reduced the burden on medical staffs by determining 37 kinds of attributes based on the outer cases of medicines collected from seven pharmacies. We constructed a database on the outer cases of medicines and analysis of it provided useful knowledge.

Keywords: Outer case, openness, medicine.

1 Introduction

In medical settings, many problems burden medical staffs. For example, they must learn how to open the outer case of each kind of medicine. However, there are many different kinds of medicines, and the outer cases of each have their own shape and specification. The number of kinds of outer packages of medicines is almost the same as the number of the medicines. Medical staffs must learn how to open the outer cases of medicine. The fragments generated by opening cases may injure fingers. So we studied how to open the outer cases of medicines to reduce such inconvenience for medical staffs. We collected the disposed outer cases of medicines from pharmacies and constructed a database, which we analyzed to get useful knowledge related to their usability.

2 Construction of a Database on Outer Case of Medicines

At present, since no database exists about the outer cases of medicines, we decided to construct a new database on the outer case of medicines. First, we collected outer cases of medicines discarded from seven pharmacies, examined them, and determined 37 attributes. Table 1 lists the attributes and explanations. These attributes were classified into the following four kinds.

a. Related to information on the surface of the outer case (11items)
b. Related to ease of opening (9items)
c. Related to capability to reseal and disposability (10items)
d. Related to actual way of opening (7items)

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 226–231, 2011.

Based on those attributes, we constructed a database of 1038 records.

Table 1. Attribute names and explanations

	Names	Description
Attributes related to information on outer case surface	JAN code	Japanese Article Number Code
	Name of medicine	Name of medicine
	Manufacturer	Name of manufacturer
	Actual producer	Name of producer
	Distributor	Name of distributor
	Dosage form	Dosage form of medicine
	Bulk	Bulk of medicine
	Surface of JAN code	Surface where JAN code is put
	Surface of GS-1data bar-code	Surface where GS-1 data bar-code is put
	Surface of expiration date	Surface where expiration date is put
	Type of GS-1 data bar-code	Type of GS-1 data bar-code
Attributes related to ease of opening	Suggested way of opening	How to open, suggested by pharmaceutical company
	Number of steps	Steps needed to open
	Sign to show place to open	With or without a sign to show place to open
	Arrow to show place to open	With or without a arrow to show place to open
	With or without fragment of paper	With or without a fragment when opened in suggested way
	Surface of opening	Surface with a place to open
	Side opened	Surface in which a cases are opened in suggested way
	Shape to be pushed	Point to push
	Versatility on dominant arm	With or without versatility on dominant arm

Table 1. (*continued*)

Attributes related to capability to reseal and disposability	Capability to reseal	With or without artifice to reseal a case
	Artifice to reseal	Artifice to reseal case
	Disposability	With or without artifice to dispose of case
	Artifice for disposal	Type of artifice for case disposal
	Prevention of tampering	With or without an artifice to determine whether case is opened or unopened
	Length	Length of outer case (mm)
	Width	Width of outer case (mm)
	Depth	Depth of outer case (mm)
	Three pictures	1) entire outer case, 2) closeup of start to open, 3) side with JAN code
	Remarks	Special instructions
Attributes related to actual way of opening	Pharmacy ID	ID's to distinguish different pharmacies
	Sequential number	Sequential number of each outer case for each pharmacy ID
	Agreement or disagreement	Agreement or disagreement between suggested way of opening and actual way
	Partly condition	
	Actual surface	Surface actually used for opening
	Condition of actual surface	Shape of side actually opened
	Actual way of opening	Actual way of opening a outer case

3 Analysis of Database on Outer Cases of Medicines

For each outer case, we compared the actual way of opening with the manufacturer's suggested way. Table 2 shows the numbers of agreement and disagreement between the actual and suggested ways. Except pharmacy A, in six pharmacies the rates of agreement exceeded 80%.

Table 3 lists the cross tabulations of the suggested ways of opening and the numbers of agreement and disagreement, except pharmacy A. It is clear that only the zipper-type has a high ratio of disagreement. Fig 1 shows examples of the actual ways of opening with the zipper-type.

We also analyzed the relations between the numbers of agreement and disagreement and the sizes of the outer cases. Fig 2 shows the analysis results. TwoStep clustering analysis was applied, and the results are as follows.

- Bigger cases tend to have a higher ratio of agreement.
- Width and depth are important for classifying clusters.
- There was no clear tendency for the smaller cases.

Table 2. Agreement or disagreement between suggested way of opening and actual way of open every pharmacy

Pharmacy ID	A	B	C	D	E	F	G
Agreement	24	130	7	187	216	133	88
Disagreement	158	24	0	7	25	9	19
Total	182	154	7	194	241	142	107
Rates of agreement (%)	13.2	84.4	100	96.4	89.6	93.7	82.2

Table 3. Cross tabulation of suggested way of opening and agreement or disagreement except pharmacy A

Suggested way of opening	Agreement or Disagreement		Total
	Agreement	Disagreement	
Push	499	19	518
Zipper	114	59	173
Pull	57	1	58
Push-pull	37	2	39
Tape	41	3	44
Others	10	0	10
Total	759	84	843

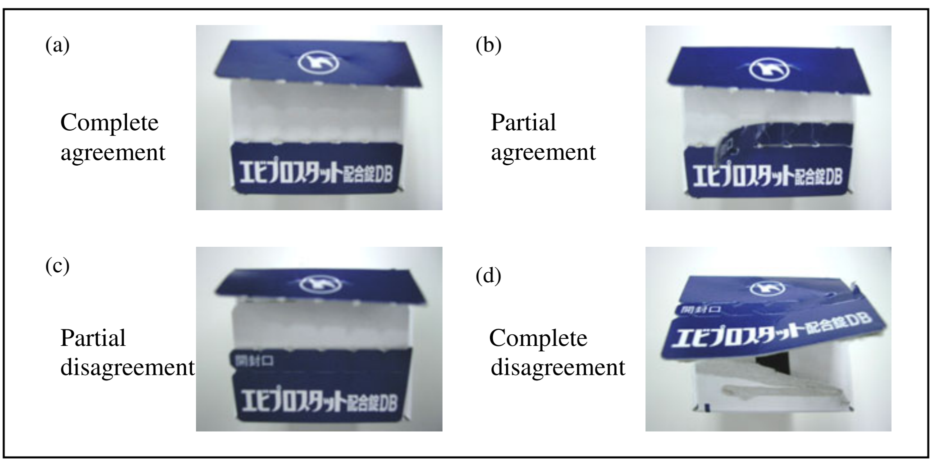


Fig. 1. Examples of actual ways of opening zipper-type

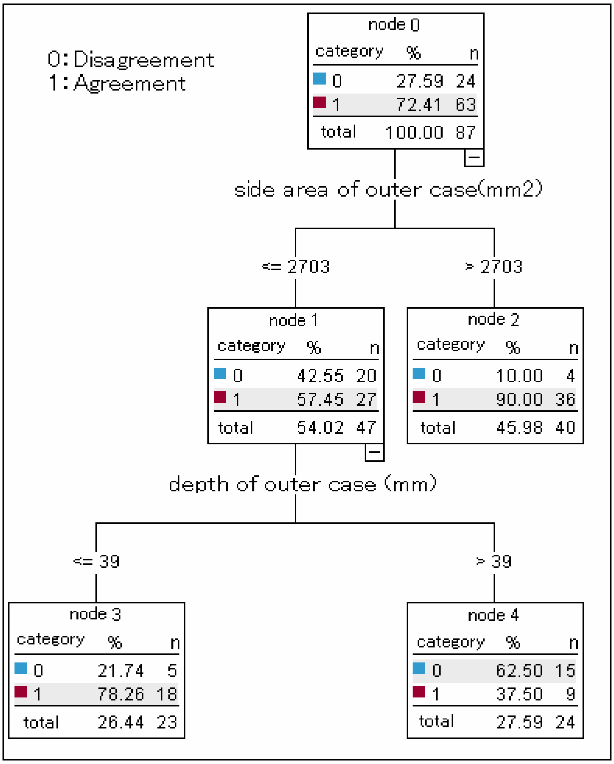


Fig. 2. Binary tree by C5.0 algorithm

4 Discussion

The high ratio of agreement between the actual way of opening and the suggested ways of opening at six pharmacies indicated that at most pharmacies, many outer cases were opened as suggested by the pharmaceutical companies. The relatively high ratio of disagreement of the zipper-type cases suggests that problems may exist in those cases.

The analysis results by the C5.0 algorithm and TwoStep clustering suggest a relationship between the size of the outer case and ease of opening them. However, the ratio of agreement may be influenced by other factors.

5 Conclusion

In medical settings, many problems burden medical staffs. For example, they must learn how to open the outer cases of each kind of medicine. To reduce such inconvenience, we determined the attributes of outer cases and constructed a database with them. Analysis results confirmed that the ways of opening differ between pharmacies. There were also problems in the outer cases of the zipper-type. In addition, the sizes of the cases are strongly related to the outer cases of medicines. However, there may be factors other than size for agreement.

Acknowledgements. We express gratitude to the medical staffs of the pharmacies for their cooperation.

Part III

Business and Commerce

Are MIS Students Learning What They Need to Land a Job?

Andrea Everard[1], Brian M. Jones[2], and Scott McCoy[3]

[1] University of Delaware, Lerner College of Business and Economics
Newark, DE 19716, USA
aeverard@udel.edu
[2] Tennessee Technological University, College of Business
Cookeville, TN 38505, USA
bjones@tntech.edu
[3] The College of William and Mary, Mason School of Business
Williamsburg, VA 23187, USA
scott.mccoy@mason.wm.edu

Abstract. The issue of declining IS enrollments is so troublesome that even ACM/AIS has redesigned its recommended curriculum in an attempt to reverse this trend [1]. Although all majors in most accredited Business Schools are required to take the Introduction to IS course, the real or perceived value of this course is unclear. This paper looks at the real or perceived value of the content of the Introduction to Information Systems course and the possible impact on declining enrollments.

Keywords: IS Curriculum, Foundation, Content.

1 Introduction

Since 2001 enrollment in MIS programs has been declining. The declining enrollments are so widespread that ACM/AIS [1] has in fact redesigned its curriculum in the hopes of guiding university programs and reversing the trend [3]. Although choice of major is based on many factors [4] [5] [6], all business majors are typically required to take the Introduction to IS course; however what real or perceived value this course provides is unclear.

2 Stakeholders

It is important that the content of the Intro to IS class is valued by those who are enrolled in it, those who are teaching it, and those who will ultimately place a value on it by recruiting business graduates. It is also critical that non-IS faculty value the Intro to IS course since information systems affect all students.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 235–236, 2011.

3 The Study

As universities chase student dollars to stay afloat in these turbulent economic times, some are focusing heavily on the student experience. Although most would agree that the student experience is important, some wonder if we are doing long-term harm to students by adjusting course content to make them happy in the short-term while making them far less marketable in the long-run. In the case of IS majors/minors, recent comments by students lead many to believe that the content of the Introduction to IS course content may not be fully appreciated by the student stakeholder [2]. And if this is true, it could be a leading factor in decisions to choose other majors/minors. Before making drastic curriculum changes, however, this study examines the curriculum content of the IS course and seek opinions from various stakeholders.

4 Summary and Conclusion

Using the topics and subtopics from the AIS curriculum, this study utilizes a web-based survey instrument to collect data from the faculty, recruiters, and students. These results will be beneficial to all stakeholders. Although we hope all three groups are well aligned, if this is not the case, the results could provide insight into where curriculum changes should be made.

References

[1] Association of Information Systems (2011), `http://enrollments.aisnet.org/Surveys%20of%20Stakeholders.ashx`

[2] Downey, J.P., McGaughey, R., Roach, D.: MIS versus Computer Science: An Empirical Comparison of the Influences on the Students' Choice of Major. Journal of Information Systems Education 20(3), 357–368 (2009)

[3] IS 2010 Final Submission, AIS website, `http://blogsandwikis.bentley.edu/iscurriculum/index.php/IS_2010_final_submission`

[4] Kim, D., Markham, F.S., Cangelosi, J.D.: Why Students Pursue the Business Degree: A Comparison of Business Majors Across Universities. Journal of Education for Business 78(1), 28–32 (2002)

[5] Noel, M.N., Michaels, C., Levas, M.G.: The Relationship of Personality Traits and Self-Monitoring Behavior to Choice of Business Major. Journal of Education for Business 78(3), 153–157 (2003)

[6] Strasser, S.E., Ozgur, C., Schroeder, D.L.: Selecting a Business Major: An Analysis of Criteria and Choice Using the Analytical Hierarchy Process. Mid-America Journal of Business 17(2), 47–56 (2002)

Promotion Project for Communication between Artisans and Consumers Supported by Media Technology

Ritsuko Izuhara, Sho Yokokawa, and Shinya Suzuki

Kanazawa Institue of Technology
3-1 Yatsukaho, Hakusan-shi, Ishikawa 924-0838, Japan
izuhara@neptune.kanazawa-it.ac.jp

Abstract. In many regions of Japan, local artisans continue to make traditional handicrafts using natural materials. However, rapid decreases in consumer demand for these products and the lack of younger successors to continue the craft have become serious problems. Possible explanations for this situation include recent changes in the Japanese lifestyle and gaps between traditional crafts made by artisans and the types of products needed by consumers. To solve these problems, the present study proposes a system to connect artisans and consumers directly. For this purpose, we established a website that facilitates communication between artisans and consumers and the generation of new ideas for craft products. We also launched a promotion event including exhibits using augmented reality (AR) and projection mapping technology to raise awareness of the project among consumers.

Keywords: Craft, Promotion, AR, Projection Mapping, Local Activation.

1 Introduction

In many regions of Japan, local artisans continue to make traditional handicrafts using natural materials. However, the traditional craft industry has come up against serious problems in recent years. For instance, consumer demand for traditional craft products has decreased rapidly, and as the younger generations move further away from traditional modes of expression and production, artisans have fewer and fewer young successors to continue their craft. Both artisans and local governments are now seeking solutions for these problems.

Kanazawa City, a "City of Hand Craftsmanship Work" located in the northwestern part of Japan, has a distinctive local culture and was accredited as a UNESCO Creative City of Craft and Folk Art in June 2009 [1]. However, even in Kanazawa, which is famous for its handicrafts, the future of traditional local craftsmanship is facing serious problems.

Observation of the situation of the traditional craft industry in Kanazawa in recent years reveals that the output of traditional craft products reached a peak in 1991 and began to decrease rapidly in 1999. One problem in business management for producers is that the number of orders from consumers has decreased [2]. More

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 237–244, 2011.

fundamentally, changes in the Japanese lifestyle have created gaps between the traditional crafts made by artisans and products needed by consumers. Due to Western influences as well as the increased emphasis on speed, usefulness, and efficiency in determining value during periods of high economic growth, people now tend to regard traditional crafts as unsuitable for their living environments.

Before we proceed, the definition of crafts should be clarified. Soetsu Yanagi, the founder of the *mingei* (folk craft) movement in Japan in the late 1920s, defined the category of figurative arts, which includes craft arts, as shown in Fig. 1. Within the category of craft arts, Yanagi defined crafts intended for practical use as “handicrafts for ordinary people”, which is one of the original meanings of “craft”[3]. However, in recent years, more value has come to be attached to appreciated crafts such as aristocratic crafts than to practical crafts. The reason for this shift is considered to be the distinction between traditional handicrafts and mass-produced goods. In this study, we follow Yanagi’s original definition and aim to promote practical crafts for ordinary people.

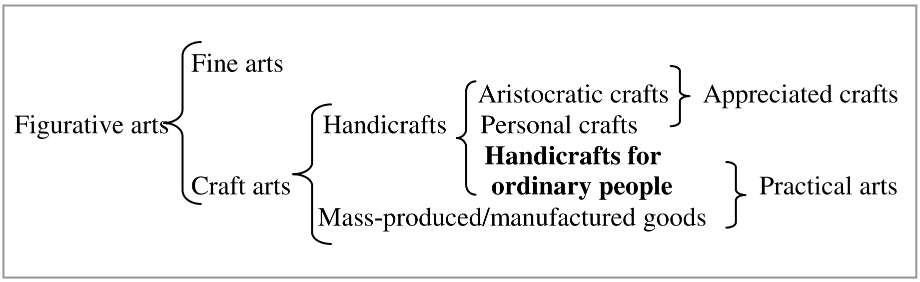


Fig. 1. Categories of figurative arts. Source: S. Yanagi.

To solve the problems currently facing the traditional craft industry, innovative products are needed. However, artisans lack adequate knowledge about the development of new products, and the cost of making trial pieces for new products is high, which prevents artisans from developing and testing new products [2].

The “Innovator Training Unit for Ishikawa Traditional Crafts” program carried out by JAIST is an activity with goals similar to those of the present study. JAIST has established a new education program to teach artisans the marketing methods necessary to develop new products [4]. After completing the program, several artisans attempted to create and market new products, such as USB memory drives made of Kutani ware porcelain. However, in doing so, the artisans had to risk negative evaluations from consumers, as they could not know consumers’ evaluations before selling the newly developed products.

In order to provide artisans with more information about consumers’ opinions during the development process, we focused on facilitating communication between artisans and consumers. The present study proposes a system that connects artisans and consumers directly and promotes the generation of new ideas for craft products by providing a forum for the exchange of opinions between community members and artisans.

2 Fieldwork and Interviews with Artisans

In order to observe the current state of products being sold in Kanazawa, fieldwork was conducted in shops selling traditional crafts. Interviews were conducted with artisans of Kaga Yuzen, a traditional technique of dyeing fabrics for kimono. The artisans discussed the marketing process of Kaga Yuzen and associated problems. To summarize briefly, wholesale dealers act as intermediaries between producers (artisans) and consumers. Dealers control all marketing flows, including planning and ordering. Consequently, artisans never meet the customers who ultimately use their products, and they cannot receive evaluations directly from customers. In order to improve their products, artisans need to hear customers' opinions; however, artisans do not have an effective way of obtaining this information. In addition, although artisans are interested in the development of new products, they frequently experience difficulty in generating ideas and strategies for product development.

3 A Community Website as a Bridge between Artisans and Consumers

We propose the use of a website to promote communication between artisans and consumers and to generate new ideas for craft products. The proposed website represents an open system in which consumers and artisans can develop new ideas for new craft products together through active dialogue. The flow of dialogue is shown in Fig. 2.

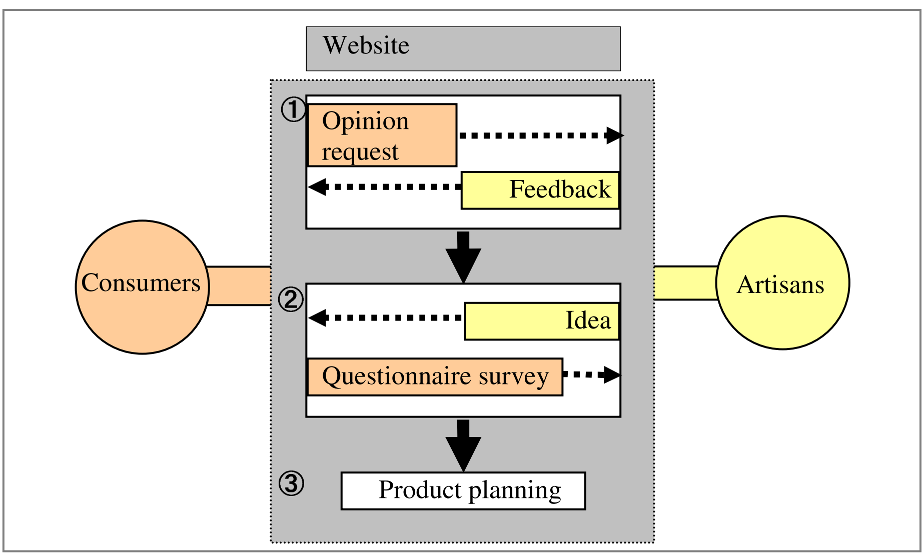


Fig. 2. Dialogue flow between consumers and artisans on the proposed website

In the first step, unspecified consumers post their opinions and needs regarding traditional craft products. In response, artisans send comments about the consumers' opinions and requests. This dialogue is repeated throughout the website. To initiate this step, a catalyst is required to draw out consumers' opinions. As catalysts, we prepared several themes for discussion and contribution, including practical products such as accessories, clothing and sporting goods. One theme, for example, requested unique ideas for combining modern practical products with traditional crafts; one of the proposed ideas was a snowboard decorated with traditional Kaga Yuzen patterns. The user who suggested this idea commented that a Kaga Yuzen snowboard would be a good representation of the identity of snowboarders in Kanazawa. In this way, the website allows artisans to obtain inspiration for new ideas for products that reflect community members' needs.

In the second step, artisans show their ideas to consumers, and consumers can send feedback to artisans in response. Furthermore, a questionnaire survey is sent to consumers asking them to evaluate the new ideas proposed by artisans. If a new idea receives a favorable review from consumers, a new product planning is formed. On the other hand, when an idea is not favorably evaluated, product planning comes to an end. In this way, the evaluation of planning before production begins is a strict but efficient and effective method of product development.

4 Website Promotion Event Using Media Technology

We held a promotion event on December 1-2, 2010 in the entrance hall of the Kanazawa City Office. This event aimed to increase awareness of traditional crafts among younger consumers and to promote our project. The concept behind this event was "Your voice and artisans' skills create new traditional crafts". In order to represent this concept, we presented ideas for new crafts generated by combining traditional crafts with suggestions from ordinary people. As the project was targeted mainly at the younger generation, the event included interactive exhibits using augmented reality (AR) and projection mapping technology to catch the attention of passers-by and to give a prominent impression of the project (Fig. 3). We also distributed flyers with our web address.

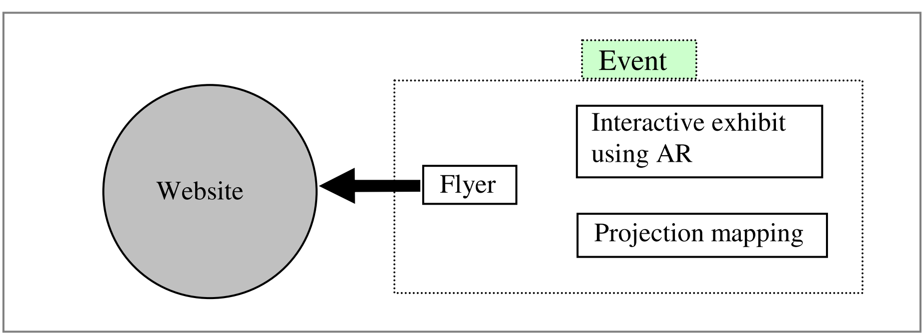


Fig. 3. Promotion event flow

4.1 New Ideas for Crafts Presented Using AR

At the promotion event, we aimed to introduce our project to visitors in order to generate new traditional crafts that reflect the voices of ordinary people, especially younger community members. In order to achieve this, we presented new ideas for crafts generated by combining traditional crafts with popular practical products. Images of these ideas were created using computer graphics (CG) and were displayed on a monitor using AR technology to show realistically shaped objects. We used ARToolKit, the open source software developed by Kato [5].

At the event, four triangular markers representing the four traditional crafts of Kaga Yuzen (dyeing), Kaga Shikki (lacquerware), Kaga Zougan (metal inlay) and Kaga Shishu (needlework) were prepared. Furthermore, four more triangular markers representing the four popular practical products of aroma goods, key cases, tableware and accessories were also prepared. Visitors were asked to choose one marker from each group. By fitting the two triangular markers together, visitors created a square marker (Fig. 4), which represented a new idea that combined one traditional craft with one popular product.

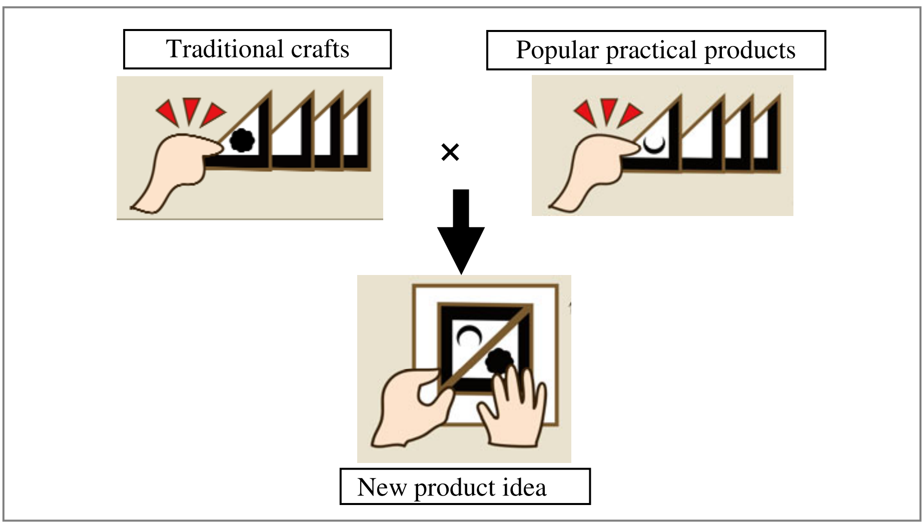


Fig. 4. Combining AR markers to form new product ideas

The combination of an aroma ball with the Kaga Shishu needlework technique is displayed on the left side of Figure 5. By fitting together the two triangular markers and placing the resulting square marker in front of the camera, a new product combining elements of each of the two triangular markers is displayed on the screen. The right side of Figure 5 shows other new ideas, including a tumbler with Kaga Yuzen patterns, a key case made with Kaga Shikki lacquer, and a bracelet with Kaga Zougan inlay.

Usually, producing a prototype is expensive and time consuming. However, using CG and AR technology allows product planning to proceed easily and at lower cost.

Fig. 5. Interactive AR exhibit showing new craft ideas

4.2 Simulation of New Product Ideas with Projection Mapping

In another exhibit at the event, projection mapping was used to simulate new ideas for traditional crafts. Projection mapping is a technique in which a two-dimensional image is projected onto a three-dimensional object. The projected image is fitted onto the object and viewers perceive the projected image as real. The display configuration of this exhibit is shown in Figure 6. In order to create the projected image, a scene was rendered using CG in which the desired objects were arranged in the same manner as the real three-dimensional objects.

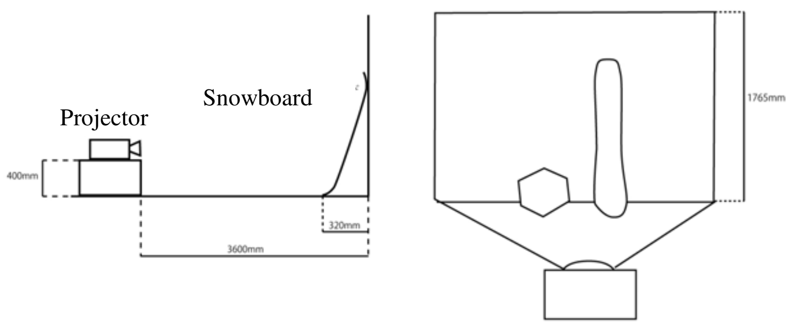


Fig. 6. Display configuration for projection mapping

We used projection mapping to display the idea of the Kaga Yuzen snowboard (Fig. 7). We created a scene with a snowboard using 10 patterns created by Kaga Yuzen artisans. The CG images were then projected onto a real snowboard in sequence.

Using this simulation method has two advantages: first, participants are able to view full-scale prototypes of new products; and second, artisans can make experimental mock-ups easily.

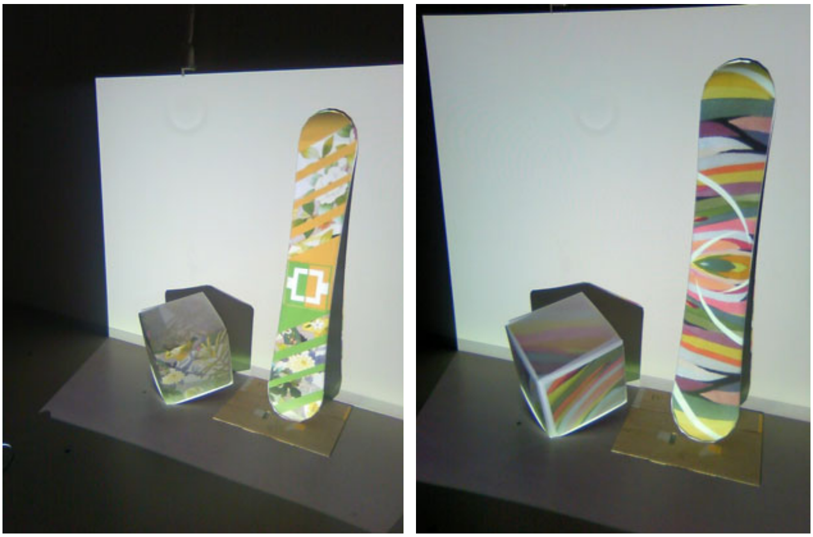


Fig. 7. Projection mapping of Kaga Yuzen snowboard

At the event, we also administered a questionnaire survey. Sixty percent of participants were in their 20s and 30s. As a result, almost all of the participants were able to understand the purpose of the project, and most of them found the AR and projection mapping exhibits interesting. In particular, we asked participants to evaluate projection mapping of the Kaga Yuzen snowboard. One participant responded that the snowboard could indicate the regional characteristics of a snowboarder in ski sites. Although the design is based on traditional Kaga Yuzen patterns, the product design seems novel because the patterns are applied to a modern sporting product.

5 Conclusion

In the present study, a website with the aim of generating new ideas for craft products through the exchange of opinions between consumers and artisans was proposed and a promotion event using AR and projection mapping technology to raise awareness of the project was reported.

After the promotion event, we received several inquiries directly from artisans who wanted to know community members' opinions through the website in order to expand their work fields. However artisans were worried about copyright issues, such as other artisans stealing people's ideas on the website. Artisans' comments offered insight into issues such as which opinions are useful and whether ideas should be shared in an open system or whether it is more important to maintain originality by keeping the idea-generation process under closer control.

We hope to proceed with our project in the future in order to continue to help artisans to adapt to the changing demands of modern society.

Acknowledgments. I would like to thank Masato Zenke and Satoshi Nishimura, staff of the City of Kanazawa, for their help in our projects.

References

1. Kanazawa Creative City Promotion Committee, Kanazawa Creative City Promotion Program 1 (2010)
2. City of Kanazawa, Action Plan of Kanazawa Traditional Craft Industry, 8–15 (2010)
3. Yanagi, S.: Kougei Bunka. Iwanami Bunko, pp. 26–29 (1985)
4. Ogata, S., Nakamori, Y., Kobayashi, T.: MOT Educational Program for Traditional Craft in Ishikawa. The Japan Society for Science Policy and Research Management, 502–505 (2007)
5. Kato, H.: ARToolKit: Library for Vision-based Augmented Reality. IEICE, PRMU, 79–86 (2002)

Why Virtual Job Recruitment Is Not Well Accepted by Generation Y?—A Case Study on Second Life

Eleanor Loiacono[1], Soussan Djamasbi[1], Bengisu Tulu[1], and Oleg Pavlov[2]

[1] School of Business, Worcester Polytechnic Institute, 100 Institute Road, Massachusetts, USA
[2] Department of Social Science, Worcester Polytechnic Institute, 100 Institute Road, Massachusetts, USA
eloiacon@wpi.edu

Abstract. Generation Y (Gen Y) forms a sizeable workforce in today's economy. Because this generation is tech savvy, it is likely that virtual worlds would serve as an ideal medium for recruiting such employees. Our research, however, showed that the participants in our study, who were part of the Generation Y population, had very little prior knowledge about recruiting in virtual worlds. Further, very few of them were familiar with Second Life, which is the major virtual world for recruitment. Even after being introduced to and spending some time looking for job fairs and job leads in Second Life, many of the participants expressed high levels of skepticism about the effectiveness of a virtual job search. Thus, Second Life is not well-known among Gen Y. There is an overall lack of awareness and thus virtual worlds are not perceived by Gen Y to be a suitable medium for job recruitment.

Keywords: Virtual Worlds, Second Life, Employee Recruitment, Generation Y, Virtual Job Fairs.

1 Introduction

The Internet and the many new technologies it affords have transformed traditional business models. Today it is difficult to imagine a world without e-mailing, online shopping, news blogging, online marketing or peer-to-peer music sharing. With the inception of virtual worlds, computer generated environments in which users participate by the means of avatars, it is quite possible that we are now witnessing the next evolutionary change of online business models [1][2][3]. Through virtual worlds, companies can potentially tap into unique talent across the globe, while keeping their costs down.

Traditional companies, encountering difficult economic times and wishing to be green, are interested in how virtual worlds can enhance their business [4]. For example, many companies have expanded their channel of communications to virtual worlds for important activities such as public outreach [5], training, and job recruitment to name a few [6]. One key factor in the success of such expansion is that the virtual world and the service that it offers are accepted by the intended target

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 245–254, 2011.

audience. In this paper, we examine one such virtual world (Second Life), one service (job recruitment), and one specific target audience (Generation Y).

We focused on Generation Y as the target audience for the recruiting efforts through Second Life. This generation, which is the largest generation in the past 50 years, is an important segment of the workforce for organizations [7]. Generation Y, also simply referred to as Gen Y, consists of those 18 to 32 years of age [8] and is the second largest demographic group (with 82 million people) after the Baby Boomer generation [9][10]. Because this formidable population is tech savvy and is expected to "push technology to new levels" [7, p.2], virtual worlds may particularly serve as a suitable channel of communication for this generation.

Because Second Life is one of the most popular virtual worlds (Virtual Worlds Review, 2008c) it served as a suitable medium for the purpose of our study. The results of our exploratory study using 29 subjects showed that besides World of Warcraft, which does not have business participants or an actual currency, Second Life was by far the most well-known virtual world (7 out of 29 or 24.14%). Additionally, Second Life is already being utilized by many businesses for recruitment [11].

In order to investigate the research question, is virtual job recruitment acceptable to Gen Y, we began by looking into the background of virtual worlds. Next, we conducted studies 1, 2a, and 2b. The results of these studies are presented in the results section of this paper and their meaning highlighted in the discussion section. Finally, our conclusions outline how companies may better harness the potential of the virtual world environment for recruiting Gen Y.

2 Background

2.1 Virtual Worlds

Virtual worlds are persistent computer-generated environments that simulate physical spaces and users participate and interact through digital actors which they control. The idea of virtual worlds has been around for some time and, currently, there are several environments that can be classified as virtual worlds. Second Life, There, Kaneva and Active Worlds are just a few. In this study we focus on Second Life, which is one of the most popular virtual worlds and is especially of interest to many businesses.

Second Life, created by Linden Labs, is a multi-user virtual environments (MUVEs) with approximately 15 million inhabitants and a user-to-user economy predicted to reach $450 million by the end of 2009 [12]. Further, Second Life is one of the best examples of a virtual world where social and economic interactions are the main drivers [13]. Monthly resident transactions for September 2009 added up to more than 3 million Linden dollars. Given the growing size of virtual worlds in general and Second Life's ability to provide a virtual world where making real money is possible, businesses and entrepreneurs around the world have become interested in gaining a presence in Second Life. Large organizations from a variety of industries, such as IBM, Cisco, Coca Cola, BMW, Adidas, Reuters, Sears, Intel, have already established their presence in Second Life (Figure 1). Many have done so as part of their recruitment process.

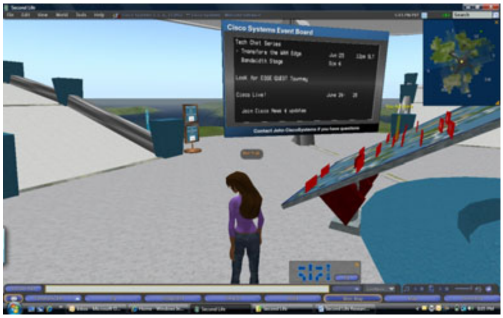


Fig. 1. Screenshot of Avatar visiting Cisco Headquarters in Second Life

2.2 Recruitment Process

Typically, recruitment includes “organizational practices and policies” developed to encourage applications, sustain a good candidate pool, and motivate the best candidates to accept job offers [14, p. 72]. The first step in the process is signaling. Here the company announces the availability of a position. This is followed by job applicants interested in the position submitting material. In the third step, information is exchanged between the applicants and company. This is often referred to as the selection process where particulars about the job are conveyed to the candidates and additional candidate information is communicated to the company. In the fourth and final step, a company makes a job offer to the appropriate candidate and is either accepted or declined. Applicant perceptions of the source of recruitment may determine whether or not a job seeker responds to a job ad or accepts a job [14]. In fact, ‘negative recruitment experiences…[are] enough to completely eliminate the organization from further consideration’ [15, p.515] and therefore applicant perceptions are important to use in evaluating the recruitment process” [14, p.83].

The traditional sources of job applications are employee referrals, print and radio ads, college campus recruitment and job hunters. The Internet, however, has become an important element in the recruiting process [16]. The source of recruitment will also be influenced by the experience needed, location, job type, etc. [14]. Support or manufacturing jobs are typically advertised locally, however, other jobs call for national or even global advertising.

Applicant perceptions of the recruitment process may determine whether or not a job seeker responds to a job ad or accepts a job [14]. Given that virtual worlds enable organizations to reach out to a new group of users that are not restricted by physical boundaries; companies can be considered as a potential recruitment venue. The three studies in this paper measured the reaction of Gen Y toward the recruiting in a Virtual World.

2.3 Recruitment in Virtual Worlds

As mentioned earlier, several organizations have begun to utilize virtual worlds for recruitment. Virtual worlds are thought to enhance the first three stages of the recruitment process. They allow recruiters to efficiently and inexpensively signal job openings to new talent (Stage 1) and gather (Stage 2) and exchange information from prospective candidates (Stage 3). Specifically, virtual worlds have allowed the collection of international applications and onboarding (an HR term for preparing a new hire for his/her job responsibilities) [4]. The U.S. Air Force is using MyBase, a multipurpose virtual world in Second Life, for recruiting civilians [5]. KPMG hosted a successful virtual job fair that had 20,000 attendees and resulted in nearly 10,000 job applications for offices in 40 countries [11]. Accenture has used its own island for virtual recruiting events as a supplement to regular job fairs and reports that it has recouped the cost of the initial investment [17]. The advantages of the virtual event are broader reach, lower cost, and standardization of events for across global offices [17]. A number of other large companies, such as EMC^2, GE, U.S. Cellular, eBay, HP, Microsoft, Sodexho, T-Mobile, and Verizon have participated in Second Life job fairs as well.

In hard economic times, such as the one we are experiencing today, new technologies that allow decreases in costs and increases in return on investment (ROI) are even more critical to a company's competitiveness and survival. Virtual worlds, including Second Life, are such technologies and are being pointed to as the next Web-like revolution [1][2]. Further, some tout virtual world environments as a good future medium for recruitment [18]. As one observer points out "companies can leverage virtual recruitment to significantly enhance their ROI in Real Life" (http://talesfromthedigitalside.com/, 2009).

Attracting job seekers to virtual recruiting events is a fundamental step in creating a successful recruitment effort. While it is argued that the use of interactive technology in business and daily activities is of particular interest to the tech savvy Generation Y [7], anecdotal evidence suggests that the soon-to-be college graduates, one of the groups companies are actively targeting through virtual worlds, may not be aware of the possibility of job recruitment to search for jobs in virtual worlds and/or such events on virtual worlds are not easy to use. For example, according to anecdotal evidence, "[v]irtual job fairs and islands of employment are not well-known…and…[e]ven if a non-technical person did find a job fair and decide to participate, there is the challenge of operating within a virtual world, such as Second Life. [Since] it takes time to become adept at controlling your avatar and getting the right appearance for an interview" [19]. Despite this anecdotal evidence little work has been done to examien this possibility directly. Hence, in this research we examine the adoption process of Second Life as recruitment tool by Generation Y. To do so we use Rogers' Theory of Innovation Adoption.

2.4 Innovation Adoption Process

According to Rogers, [20] all innovations (in our case Second Life as a recruitment tool) follow a five stage process before becoming fully adopted. In stage 1, *knowledge*, a person becomes aware that the technology exists. However, s/he usually lacks detailed information on the innovation. Further, s/he has not yet been "inspired" to gather more information about it. In stage 2, *persuasion*, a person learns how the technology works. The person's interest in the innovation is peaked and s/he actively seeks information about it. It is common for people during this stage to experience frustration and a lack of confidence in the innovation. In stage 3, *decision*, a person starts to consider the advantages and disadvantages of the innovation before determining whether to adopt or reject it. S/he struggles to determine if using the innovation would actually be advantageous in her/his life. The process may require seeking advice and assistance from trusted sources before moving to the next stage.

In stage 4, *implementation*, a user gains a sense of confidence in using the innovation. S/he is not currently committed to the innovation and its continued use. S/he is, in a sense, "test driving" the innovation. An evaluation of the ease of using and interacting with the innovation is being made. If the innovation is not easy to learn and use, a person is likely to reject it. In stage 5, *confirmation*, a user makes a final decision to adopt the innovation. There is no more consideration of the value of the innovation. The user has determined it is valuable and incorporates it into her/his work or personal life environment [20].

3 The Studies

As suggested above, any successful innovation must proceed through the five stages of adoption. This research looks at virtual worlds as an innovation and, in particular, their use as a medium for job seeking by Gen Y. To investigate this issue more directly and rigorously, we examine the current stage of Second Life as a job recruitment tool by Generation Y in the innovation adoption process and the factors that are likely to motivate Generation Y to move from its current stage to the next stage in the process. In particular, we conducted three exploratory studies which are discussed in the following sections.

Study 1: The objective of the first study was to test Gen Y's position in regards to the first step in the innovation adoption process, namely knowledge. Forty-five Gen Y subjects responded to questions regarding their familiarity with virtual worlds and in particular Second Life. All subjects were between the ages of 18 and 28 years of age. Over half of the subjects (60%) were male. The subject pool was drawn from technically fluent college students attending a northeastern university. In return for their participation, subjects were given extra credit in a course. Data was collected using an online questionnaire which focused on their current knowledge and use of virtual worlds, such as Second Life. *Results*: After analyzing the data, it became clear that, although technically fluent, Gen Y users are not very familiar with Second Life. Though Second Life is the most prominent online virtual world used by companies to recruit Gen Y users, this target audience has little to no knowledge that it exists, let alone would provides possible job opportunities. Out of the 45 subjects included in the study, only three (7%) were familiar with Second Life. These results, which confirm prior anecdotal evidence, suggest that the participants in this study were in the first stage of the innovation adoption process.

Study 2a: According to Roger's model an individual can be persuaded to use a technology if learning about the technology peaks his/her interest. Thus, we conducted a second study, which tested whether knowledge about job recruitment through Second Life captures users' interests. To do so we looked more specifically at job fairs in Second Life and subjects' willingness to attend, such events even if they were not familiar with virtual worlds. Included in this study were 21 subjects, who ranged in age from 18-28 (average age of 20.05 years). Of the total subjects, nineteen were males. Though multiple questions were collected, the two questions of greatest interest to this study were: 1) Have you ever attended a job fair in Second Life and 2) If you are not a current member of Second Life, would you consider joining Second Life as part of your job search process? *Results*: The results of this study confirm those of Study 1, but add additional insight. All 21 subjects replied that they were not familiar with job fairs in Second life. This is consistent with Study 1. If those in Gen Y are not familiar with Second Life in general, they are even less likely to be familiar with job fairs held within Second Life. Interesting, however, nine (43.86%) of the subjects were willing to consider joining Second Life as part of their job search process. This suggests that the prospect of increasing one's recruitment possibility is likely to persuade Gen Y to use Second Life as a recruitment tool.

Study 2b: The results of the previous study inspired a further exploratory study. The results of the first two studies show that Generation Y is in the knowledge state of the innovation adoption process. A major goal of this study was to determine if subjects, given the opportunity to try out and thus evaluate Second Life as a job recruitment tool (guiding students through knowledge to persuasion), would decide to accept it rather than reject it. Another goal was to determine possible factors that may entice our participants to adopt Second Life as a potential recruitment tool, as well as, those that may prevent its adoption. To do this, 68 Gen Y subjects were recruited to participate. The subjects came from the same pool as in Study 2a. They were college students attending a northeastern university who were tech savvy. The participants were 73.5% male and their average age was 20.9. The subjects were first asked to go to Second Life and create an account, if they did not already have one. They then created an avatar, completed the avatar training process, and finally found as many recruiting events as possible. Subjects spent an average of 14.5 minutes training their avatar in Second Life and 40.14 minutes looking for a recruiting event. The average number of recruiting events visited by subjects was 3.8.

After finishing the task, subjects' overall opinion of the virtual world (Second Life) was captured through the System Usability Scale (SUS) commonly used in industry laboratories [21]. Finally, they were asked to rate the effectiveness of several job search methods (such as job fairs, direct employee contact through websites, internet job boards, college career centers, newspaper, and classified ads) and a virtual world (Second Life). *Results*: Consistent with the results of the two previous studies, the results showed that only a few of the participants were active Second Life users. Of the 68 respondents, only 8 (11.76%) had a Second Life account. Further, of those who had a Second Life account, 62.5% could not remember the last time they visited Second Life. Interestingly our analysis showed that most of our tech savvy Gen Y participants were not using virtual worlds in general; only 17 (25.00%) of the respondents had participated in any massively multiplayer online games (MMOG), such as World of Warcraft, Final Fantasy, or EverQuest.

Participant's perception of usability of Second Life was calculated through SUS measures. As recommended (Brooks 1986) an overall SUS score was calculated. The calculated SUS score (37.94) was well below the average usability score which ranges from 65 to 70 for websites. Participants' captured comments, regarding their interactions with Second Life further corroborate the low SUS score that reflects less desirable interactions. These comments will be discussed later.

Another exploratory test of the SUS items provided interesting insight about usability of Second Life. Three measures, which get at the heart of how easy the system is to learn and use, showed that the participant did not find the system particularly easy to use. These items, which were all captured on a 7 point scale, were: "I thought Second Life was easy to use" (3.67); "I would imagine that most people would learn to use Second Life very quickly" (3.85); and "I need to learn a lot about Second Life before I could effectively use it" (4.97). Further, the SUS item particularly focused on whether a user would use Second Life in the future had an average of 2.00 out of 7.

An additional questionnaire was used to compare users' evaluation of Second Life as recruitment tool to other job search methods. This survey asked user to rate the effectiveness of several recruitment methods on a 7 point Likert scale with 1 indicating the least favorable response and 7 the most favorable response. The analysis of this data showed that the job search method thought to be most effective by a majority of subjects was the "traditional" real world job fair (5.52 out of 7 point Likert scale). It was followed by college career centers (5.21) direct employment contact through a company's website (4.94), private employment offices (4.34), internet job boards (4.14), classified ads (3.64), and Second Life recruiting (3.01). Subjects' comments indicated that they believed that "real-life job fairs allow[ed] job candidates to present their skills better than a virtual career fair would." An ANOVA test showed ($F = 35.02$, $p = 0.000$) that there was a significant difference in the ratings for the seven job search methods. A subsequent t-test (t stat = 2.95, $p = 0.004$) revealed that Second Life recruiting was rated significantly lower than classified ads in newspapers and journals and thus significantly worse than any other job search tool rated in this study.

"I only found empty job recruitment centers" and "I disliked the lack of people there." This sentiment was echoed by numerous other subjects.

Focusing on enticements that would encourage Gen Y's participation in virtual world recruiting events, subjects ranked (1) "meet current employees," (2) "hear about current work projects," and (3) "get a feel for the "real" headquarters" as the most important incentives. They did not consider give-aways or monetary rewards to be much of an incentive and ranked them as the least effective ways to induce attendance at recruiting events.

When considering whether to use Second Life as a job search tool, most subjects (60%) revealed that they would not use Second Life as a job searching tool at all and the rest said they would only use it sparingly. Comments were mixed about the value of a virtual pre-screening interview. Some felt that, "Second Life [could] be used as a very powerful tool in preparing for interviews and would let [the subject] know more about the company" and that they would be willing to attend an event." However, others felt that, "presentation is very crucial in an interview and this requires a personal presence…an avatar won't be able to portray…personality traits in an effective manner to the employer". One comment conveyed the overall sentiment of caution, yet excitement for Second Life, "I [would] need to explore a little more to see the pros and cons of the interview process in Second Life. But, using it for the first time, I found it quite interesting."

Even given these positive comments, the majority of subjects preferred face-to-face interviews over any other form of interviewing. However, if asked to participate in a prescreening interview through a virtual world, 60% agreed they would.

The most important factors in determining a company's presence in Second Life for Gen Y are related to the event itself. Of upmost importance is the quality of the event (6.21 on a scale of one to seven). Next is the frequency of events (5.37) and number of people attending events (5.16).

4 Discussion

Overall, what the results of the three studies suggest is that Gen Y is not overly enthusiastic about virtual online world recruitment. Referencing Roger's model, it appears that Gen Y is simply not very aware of virtual worlds. As mentioned earlier, only 25% of respondents had actually participated in MUVEs and less than 12% had a Second Life account. Even though Gen Y is quite aware and familiar with many technologies in general and Web-based applications specifically [7], their awareness of Second Life and other MUVEs, is quite low. An effective way to make Gen Y aware of a phenomenon is by publicizing in places that they often visits, such as social networking sites, YouTube, and concerts [10]. Thus, companies may benefit from cross-advertising their recruitment efforts in this way. They will increase awareness and their image as a tech savvy company. Although some within Gen Y are familiar with virtual worlds, indicating a move to Stage 2, learning how a virtual world works, they do not seem to be able to make that leap to Stage 3 where they can really see virtual worlds, such as Second Life, as a benefit in their job search. Although Gen Y is very motivated to complete their job search, they found maneuvering and intuitively feeling their way around Second Life to be somewhat difficult. This lack of usability is very frustrating to many and resulted in a number of respondents leaving the world frustrated.

So, what do companies looking to recruit applicants in virtual worlds need to do in order entice Gen Y applicants? Firstly, it is important that if a company makes the leap into recruitment via virtual worlds, it make a full commitment. Most of our participants mentioned the lack of "a real person" at organizational sites they visited. In addition, they found the frequency of announcements which simply directed them to static web sites to be unhelpful. In order for organizations to gain value from these new technologies, they need to first show their own commitment by having a real and strong presence. Secondly, companies can improve recruiting efforts by having avatars at their sites. The absence of such a representative is the equivalent of having a "contact us" link/form on a webpage, but no one to respond to the user's feedback. Thirdly, companies must convey to Gen Y that there is a value to using this new way of recruitment. One way to do this is to cross-connect with existing social networking sites. As many subjects stated, they would use it, "if it was easier than person-to-person or gave me more opportunities". Further, they would like to see it, "linked with current social network[s] [such as] Linked In. Finally, virtual worlds have the unique ability to increase the job pool of candidates, not only by extending their reach to different geographical location but also by making their recruiting efforts more accessible to those with various disabilities, such as those with mobility issues.

5 Future Research and Limitations

This subject pool is an appropriate group to sample for several reasons. First, college students are typically active job seekers, who are seeking either internships or full-time employment, and thus serve as a good source of evaluators of recruitment tools. In fact, over half (55.55%) of the subjects in this study were actively seeking a job at the time of the study. Secondly, college students (in particular Gen Y) tend to accept new technologies quicker than other groups. For this reason they are a good source for evaluating new technologies for job recruitment. That said, however, our subjects tend to be on the younger-side of the Generation Y. We plan to expand the pool of Gen Y to include a larger span of ages in the future. Further, in this study, we assumed that subjects, given that they were senior level college students, were in a similar stage of job search. In future studies, we plan to ask subjects specifically to identify where they are in their job search process. Finally, we examined Second Life from the perspective of job seekers, however, future work would benefit from investigating company perspectives, such as how accurately an avatar is representative of the job seeker.

6 Conclusion

Our results, however, are still very relevant and informative. Second Life, one of the most popular virtual worlds, is not necessarily well-known among Gen Y. There is an overall lack of awareness and thus virtual worlds, Second Life in particular, are not perceived by Gen Y as a suitable medium for recruiting. That said, subjects' responses did indicate that they were still open to further investigation of virtual worlds as a recruitment tool. To them, it "depends on the amount of effort to build and maintain the avatar and the quality of the possible employers." If they see virtual worlds as a benefit to their job search, they are much more likely to adopt it. Further, they are more likely to encourage their peers to use it as well. As pointed out earlier, there are several steps that companies can take to improve their recruitment within virtual worlds, but all require a true commitment by the company to follow through on its initiative.

References

1. Bainbridge, W.S.: The Scientific Research Potential of Virtual Worlds. Science 3(17), 472–476 (2007)
2. Havenstein, H.: Virtual worlds will soon be as important as Web to companies. Computer World (2008)
3. Mennecke, B.E., McNeill, D., Roche, E.M., Bray, D.A., Konsynski, B., Townsend, A.M., Lester, J.: Second Life and Other Virtual Worlds: A Roadmap for Research. Communications of the AIS 22(20), 371–388 (2008)

4. Virtual Worlds News Unisfair Sees Growth For Smaller Events and Training. Virtual Worlds News.com (2009)
5. Virtual Worlds News MyBase Launches in Second Life (December 02, 2008b)
6. Benner, K.: I got my job through Second Life. Fortune (2007)
7. Holley, J., Generation, Y.: Understanding the Trend and Planning for the Impact. In: 32nd Annual IEEE International Computer Software and Applications Conference (2008)
8. Fox, S., Jones, S.: Generations Online in 2009. PEW Internet & American Life Project (2009)
9. Djamasbi, S., Tullis, T., Siegel, M., Ng, F., Capozzo, D., Groezinger, R.: Generation Y & Web Design: Usability Testing through Eye Tracking. In: Proceedings of the Fourteenth Americas Conference on Information Systems, Toronto, Canada (2008)
10. Fields, B. Marketing to Gen Y: What You Can't Afford Not to Know. Startup Nation (2009)
11. Virtual Worlds News Quick Stat: KPMG Virtual Job Fair Draws 20k Registrants (November 18, 2008c)
12. Linden Predicts $450M In Second Life User Transactions For 2009, Virtual Goods News (2009), http://www.virtualgoodsnews.com/2009/04/linden-predicts-450m-in-second-life-user-transactions-for-2009.html
13. Hendaoui, A., Limayem, M., Thompson, C.W.: 3D Social Virtual Worlds: Research Issues and Challenges. IEEE Internet Computing 12(1), 88–92 (2008)
14. Harris, M.M., Lasson, E.D.: Recruitment. In: Edwards, J.E., Scott, J.C., Raju, N.S. (eds.) The Human Resources Program-Evaluation Handbook, pp. 71–88. Sage Publications, Thousand Oaks (2003)
15. Rynes, S.L., Bretz, R.D.J., Gerhart, B.: The importance of recruitment in job choice: A different way of looking. Personnel Psychology 44, 487–521 (1991)
16. Neil, A.: Applicant and Recruiter Reactions to New Technology in Selection: A Critical Review and Agenda for Future Research. International Journal of Selection and Assessment 11(2-3), 121–136 (2003)
17. Virtual Worlds News Accenture Recruiting in Second Life Paid for Itself (August 15, 2008a)
18. Athavaley, A.: A Job Interview You Don't Have to Show Up For. The Wall Street Journal (2007)
19. D'Angelo, D.: Try Second Life Beyond the IT Department (2008)
20. Rogers, E.M.: Diffusion of Innovation. The Free Press, New York (2003)
21. Finstad, K.: The System Usability Scale and Non-Native English Speakers. Journal of Usability Studies 1(4), 185–188 (2006)

Investigating Online Advertising in Chile

Scott McCoy[1], Cristóbal Fernández Robin[2], and José Luis Cortés [2]

[1] The College of William and Mary, Mason School of Business
Williamsburg, VA 23187, USA
scott.mccoy@mason.wm.edu
[2] Universidad Técnica Federico Santa María, Departamento de Industrias
Av. España 1680, Valparaíso, Chile

Abstract. Internet advertising continues to show signs of healthy growth despite the current economic downturn, but online advertisements are often considered undesirable by most users. In this study, we focus on the impact these online advertisements have on users in Chile. The study was conducted in a laboratory setting with 80 student subjects. Results are helpful for both researchers and practitioners.

Keywords: Online Advertising, Electronic Commerce, Attitudes, Intentions, Latin America.

1 Introduction

Online advertising continues to grow as a percentage of overall advertising despite the perceived intrusiveness and irritation on behalf of users. Much of the previous research in online advertising has focused on specific types of ads in studies conducted primarily in the US [1, 2, 3]. We attempt to build on this research in investigating the impact of online advertisements in Chile.

2 Prior Research

Research has reported that users develop negative attitudes towards ads and avoid them when possible [4], [5]. A recent study [2] found that although users remembered the online ad content when presented with website content, the attitudes toward those ads were low suggesting that users have little tolerance for online advertisements. These negative attitudes affect brand perceptions [6].

3 Proposed Research

This research focuses on previously investigated factors in the context of web users in Chile. The experiment was conducted in Chile in Latin America. The focus on this area was for several reasons. First, Latin America is one of the fastest growing areas

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 255–256, 2011.

in advertising. Second, Chile has one of the highest Internet penetration rates at over 50% [7].

The study was conducted in an experimental setting to control the location and frequency of the advertisements, as well as to allow measurement of all of the outcome variables. The use of a real website and real ads was intentional to make the experiment as realistic as possible.

4 Summary and Conclusion

The data for this project has been collected and the results will be shared at the conference. These results will guide future work for both researchers and practitioners.

References

1. McCoy, S., Everard, A., Loiacono, E.: Online Ads in Familiar and Unfamiliar Sites: Effects on Perceived Website Quality and Intentions to Reuse. Information Systems Journal (ISJ) 19(4), 437–458 (2009)
2. McCoy, S., Everard, A., Polak, P., Galletta, D.F.: The Effects of Online Advertising. Communications of the ACM (CACM) 50(3), 84–88 (2007)
3. McCoy, S., Everard, A., Polak, P., Galletta, D.F.: An Experimental Study of Antecedents and Consequences of Online Ad Intrusiveness. International Journal of Human Computer Interaction (IJHCI) 24(7), 672–699 (2008)
4. Abernethy, A.M.: Differences Between Advertising and Program Exposure for Car Radio Listening. Journal of Advertising Research 31(2), 33–43 (1991)
5. Krugman, D.M., Johnson, K.F.: Differences in the consumption of traditional broadcast and VCR movie rental. Journal of Broadcasting & Electronic Media 35, 213–232 (1991)
6. MacKenzie, S.B., Lutz, R.J.: An Empirical Examination of The Structural Antecedents Of Attitude Toward the Ad in an Advertising Pretest Context. Journal of Marketing 53(2), 48–65 (1989)
7. InternetWorldStats, Usage and Population Statistics, `http://www.internetworldstats.com/` (last accessed in January 2011)

Analysis of Customer Satisfaction on the Stiffness of Outside Panels of Passenger Cars

Ilsun Rhiu[1], Taebeum Ryu[2], Byungki Jin[1], and Myung Hwan Yun[1]

[1] Department of Industrial Engineering,
Seoul National University, Seoul, 151-744, Korea
[2] Department of Industrial and Management Engineering,
Hanbat National University, Daejeon, 305-719, Korea
1sunr@naver.com, tbryu@hanbat.ac.kr,
{bkjin01,mhy}@snu.ac.kr

Abstract. The purpose of this study is to analyze customers' satisfaction for the stiffness of outside panels of passenger cars. Including 'satisfaction', four affective variables were selected for rating affection of outside panel stiffness. About fifty customers evaluated the hood and trunk lid of nine midsize passenger cars with a developed questionnaire in the study. The stiffness of the hood and trunk lid for the nine vehicles was measured using stress-strain curves. It was found that customers were more satisfied as the slope of the stress-strain curves increased, and the decrease at a point in the curve had negative effects. In this study, the levels of satisfaction of outside panel stiffness were grouped by stress-strain curves, and it is likely that the affective quality of outside panel stiffness can be controlled by them.

Keywords: Stiffness of outside panel, Stress-strain curve, Passenger cars, Affective design, Affective quality control.

1 Introduction

The affective quality of passenger vehicles such as the look and feel as well as functional performance (e.g. horse power and fuel consumption efficiency) is becoming important factor in customers' purchase decision. Affection is defined as a customer's psychological response to the perceptual design details of the product [3]. Product designs that do not consider customers' affection may essentially be weakened [6]. In this context, the unexpected and excessive deformation of outside panel of vehicles, which occurred in customers' contact, can bring cheap feeling. Thus, stiffness of the outside panel is important for the affective quality of vehicles.

The main challenge for affective design is to accurately grasp the customer's affective needs and subsequently to design products to meet those needs. Two key research issues of affective design are (1) how to measure and analyze human reactions towards affective design; and (2) how to assess the corresponding affective design features. Regarding the first issue, the semantic deferential method (SD) [16] was used most frequently to investigate customers' perception on products. Many

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 257–265, 2011.

previous studies have used this method to study specific aspects of different product aspects, including styles, colors, and other attributes in product design [8, 2, 7, 18, 1, 12, 11]. As a qualitative approach, focus groups are utilized to provide a clear insight on the usefulness of a new product's design [10]. Similar techniques include one-on-one interviews and similarity–dissimilarity attribute rankings [4]. In reality, subjective impressions are difficult to translate into verbal descriptions, affective needs are relatively short-lasting emotional states and tend to be imprecise and ambiguous [5].

Kansei engineering has been well recognized as a technique of translating consumers' psychological feelings about a product into perceptual design attributes [14]. Nagamachi proposed six technical styles of Kansei engineering methods with applications to the automobile industry, cosmetics, house design, and sketch diagnosis [13]. Other product design practitioners in a variety of areas have utilized such computational approaches [9, 15, 21, 17, 22, 19, 20]. These studies quantified people's perceptions about the product in order to translate a consumer's perceptions into design elements.

Few studies have been conducted on the design of outside panel stiffness in terms of customer satisfaction. Many Kansei engineering studies on automobiles have focused on the visual design characteristics of interior and exterior parts [9, 14, 15]. In the study of [20], tactile feeling was considered as well as visual properties in evaluating the satisfaction of interior materials of passenger cars. But few studies exist focusing on the tactile feeling of the stiffness of outside panels.

This study attempted to analyze customers' satisfaction of outside panel stiffness of passenger cars. To do this, this study conducted following tasks: 1) a questionnaire was developed to evaluate customers' affection on the stiffness of vehicle outside panels, 2) design variables related to the stiffness of the panels were selected, 3) an experiment to evaluate customers' affection for the outside panels of various passenger cars was performed, and 4) statistical analysis was conducted based on the experiment data to analyze customers' satisfaction of outside panel stiffness.

2 Method

2.1 Design Variables of Outside Panel Stiffness

The stiffness of an automobile's outside panels was measured using the stress-strain curve (Fig. 1 as an example). From a stress-strain curve, two design variables: the slope of curve and type of decrease of the curve at a point (called canning) can be defined. The slope of curve is defined as the slope between the start and end point of a curve. There are infinite types of canning in terms of its range and shapes. It was difficult to collect all kinds of outside panels for the experiment which have various stress-strain curves in the way of factorial design. Thus the stress-strain curve for an outside panel stiffness itself was selected for design variable. The study obtained the stress-strain curve from the weakest point of an outside panel and the value of the two variables related to the stress-strain curve was taken for the further analysis.

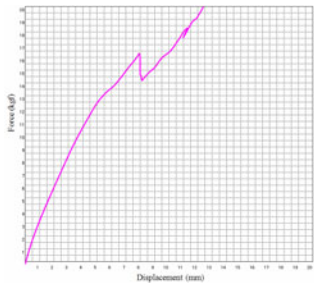


Fig. 1. Stress-strain curve example of an outside panel of passenger cars

2.2 Affective Variables

Seven initial affective variables of outside panel's stiffness were collected through Korean adjectives related to touch feel, web survey for customers' experience of contact with vehicle outside panels and expert review. After integrating the initial affective variables, four affective variables were selected based on pilot-test. Selected variables were 'satisfaction', 'hardness', 'consistency', and 'thickness'. Definitions of selected affective variables are given in Table 1.

Table 1. Selected affective variables for the stiffness of an automobile's outside panel

Affective variables	Definition
Satisfaction	Degree of satisfaction in terms of the automobile outside panel's stiffness when pressing it
Hardness	Degree of how much impact the outside panel can take when pressing it
Consistency	Degree of consistency in deformation of the automobile outside panel when pressing it
Thickness	Degree of how thick the automobile outside panel feels when pressing it

2.3 Evaluation Questionnaire

The developed questionnaire to evaluate customers' affection on the panel stiffness consisted of 1) the questions for basic information of customers, 2) explanation script of the evaluation method and target parts of vehicles, 3) the question to rate customer's affection, and 4) post-test questions. The questions for basic information of customers were included to obtain the demographic data, driving experience and the frequency of contact with the outside panel of vehicles. In the explanation script of evaluation method, a scenario to set up an evaluation context, a detailed task for evaluator to perform (pushing hard the panel with their palm several times), and the explanation of affective s selected in this study were included. In the question to rate customer' affection, the selected affective variables are rated by using 7-points SD scale. The post-test questions were included to clarify and reason customers' rating.

2.4 Outside Panel Parts and Vehicles

Two parts including the hood and trunk lid were selected to examine customer satisfaction with outside panels (Fig. 2). These selected parts are most frequently touched with the driver.

The present study used nine midsize passenger cars to measure the design variables of outside panels and customer satisfaction. The vehicles were placed at a yard of an auto manufacturing company; of the vehicles, 2 were domestic and the other foreign, having various characteristics of stiffness.

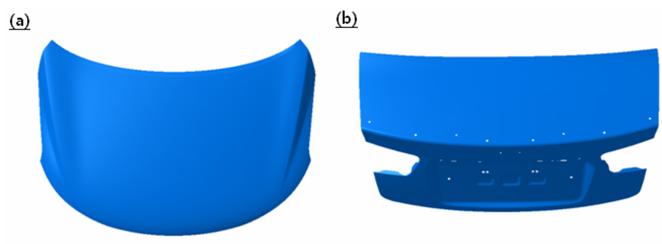


Fig. 2. Selected outside panels: (a) Hood, (b) Trunk lid

2.5 Participants and Procedure

A total of 54 males participated in the outside panel affection evaluation for the nine vehicles. Of the participants 25, 17, 8, and 4 were in their 20s, 30s, 40s and 50s respectively.

The evaluation experiment in the study consisted of three sessions: introduction, satisfaction evaluation, and debriefing. At the introduction session, the purpose and method of evaluation were explained to the participants, and the basic questions in the questionnaire were answered. Then, in the evaluation session, each participant visited the 9 vehicles and evaluated the outside panels of the 2 parts in each vehicle by following a predetermined order (the evaluation orders of the vehicles were randomized by the balanced Latin-square design to counterbalance the effects of learning and fatigue). Meanwhile, participant pushed the predefined point of the outside panels, which was marked by the experimenter and on which stress-strain curves were measured. Lastly, at the debriefing session, the post-test questions were answered.

3 Results

3.1 Analysis of Relationship between Affective variables and Design Variables

An ANOVA with mixed-factors design was conducted to analyze the effect of outside panels on affective variables. The factors involved in the experiment were type of stress-strain curve, age and their interaction. The variable of stress-strain curve type was within-subject factor and the other variable (age) was between-subjects factor in

Table 2. Summary of ANOVA results

Part	Independent Variable	Affective variables	df	*F*	*P*
Hood	Stress-strain curve	Satisfaction	8	10.46	0.0001
		Hardness	8	19.61	0.0001
		Consistency	8	19.00	0.0001
		Thickness	8	19.68	0.0001
	Age	Satisfaction	3	0.80	0.5012
		Hardness	3	0.43	0.7322
		Consistency	3	0.42	0.7388
		Thickness	3	0.07	0.9769
Trunk Lid	Stress-strain curve	Satisfaction	8	16.27	0.0001
		Hardness	8	16.29	0.0001
		Consistency	8	16.31	0.0001
		Thickness	8	15.22	0.0001
	Age	Satisfaction	3	2.28	0.0910
		Hardness	3	1.67	0.1846
		Consistency	3	1.90	0.1411
		Thickness	3	1.30	0.2832

Table 3. Result of 'satisfaction' SNK Grouping in hood

SNK Grouping		Mean	N	Stress-strain curve
	A	5.1887	53	8
	B	4.5185	54	6
	B	4.2778	54	5
C	B	4.1667	54	3
C	B	4.1296	54	9
C	B	4.0926	54	7
C	B	4.0741	54	1
C	D	3.5000	54	4
	D	3.1111	54	2

Table 4. Result of 'satisfaction' SNK Grouping in trunk lid

SNK Grouping			Mean	N	Stress-strain curve
	A		5.6111	54	1
B	A		5.4815	54	9
B	A	C	5.3148	54	8
B	A	C	5.2075	53	4
B	D	C	4.9815	54	6
	D	C	4.8519	54	3
	D	C	4.8148	54	7
	D		4.5926	54	2
	E		3.8148	54	5

the experiment. The results are presented in Table 2. All the affective variables were influenced by stress-strain curve type for both the hood and trunk lid. But the effect of age and interaction between the curve and age were not significant on all the affective variables.

Table 5. Result of Hood's Conjoint Analysis (F=46.9141, p=0.0001, R^2=0.6311)

Preference	Attribute	Relative Importance	Attribute Value	Utility
Satisfaction	Hardness	25.7994	1	-0.79392
			2	-0.37392
			3	-0.24092
			4	0.14967
			5	0.38742
			6	0.35038
			7	0.52130
	Consistency	31.7518	1	-0.82487
			2	-0.68116
			3	-0.32824
			4	0.14189
			5	0.11998
			6	0.79380
			7	0.77861
	Thickness	42.4488	1	-1.03201
			2	-0.64922
			3	-0.26271
			4	-0.11734
			5	0.24654
			6	0.68277
			7	1.13198

Table 6. Result of Trunk Lid's Conjoint Analysis (F=74.8220, p=0.0001, R^2=0.7330)

Preference	Attribute	Relative Importance	Attribute Value	Utility
Satisfaction	Hardness	47.0227	1	-1.71442
			2	-0.95498
			3	-0.11814
			4	0.33094
			5	0.59693
			6	0.83033
			7	1.02936
	Consistency	18.2051	1	-0.14561
			2	-0.42257
			3	-0.35755
			4	-0.25838
			5	0.03507
			6	0.51386
			7	0.63970
	Thickness	34.7722	1	-1.10207
			2	-0.43810
			3	-0.34539
			4	0.13849
			5	0.23078
			6	0.58940
			7	0.92689

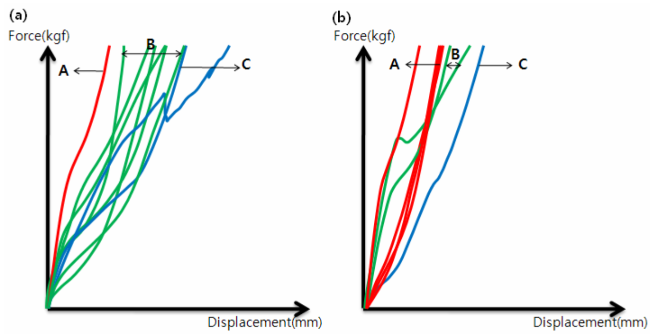


Fig. 3. Stress-strain curve of outside panel ((a) : Hood, (b) : Trunk lid)

Nine stress-strain types were grouped in terms of customer satisfaction by using SNK (Student Newman-Keuls) method. The nine stress-strain curves of the hood were grouped to the maximum of four groups (Table 3). The stress-strain curves of the trunk lid were grouped to the maximum of five groups (Table 4).

3.2 Analysis of Relationship between Affective 'Satisfaction' and Related Affective variables

Conjoint analysis was conducted to analyze the relationship among 'satisfaction' and its related affective variables. The results are presented in Table 5, 6. According to Table 5, results indicated that 'satisfaction' was most influenced by 'thickness' in hood. According to Table 6, results indicated that 'satisfaction' was most influenced by 'hardness' in trunk lid. From these results, the key affective variables could be changed by parts of vehicles' outside panel.

4 Discussion and Conclusions

As expected, it was found that the stiffness was a significant factor on customers' affection on outside panels of passenger cars, and its effect was not different by age. The collected nine stress-strain curves of the hood and those of the trunk lid for the midsize passenger cars could be grouped by the SNK results like Fig. 3. Group A was the most satisfied outside panel, and group C was the most unsatisfied outside panel. There was a tendency that the customers rated high as the slope of curve increased in both the hood and trunk lid. In addition, the canning had a negative effect on the satisfaction of the outside panel.

With the results of the study, it is likely that the affective quality of outside panel stiffness can be controlled by the stress-strain curves. Using the result of the study, the level of affective satisfaction for a hood or trunk lid stiffness can be estimated based on its stress-strain curves. And the designers of outside panels are able to know how to make the stress-strain curves of panels for the desired level of satisfaction. Likewise, the method of the experiment and analysis in the study can be extended for the other parts of vehicles' outside panels to control the affective quality of them.

In addition, the analyzed relationship between ‘satisfaction’ and its related affective variables can be used to understand the characteristics of satisfaction for vehicles’ outside panel. This study specifically explored the quantitative relationship among affective variables and between affective variables and design attributes. Based on these results, it is expected that customers’ satisfaction for stiffness of outside panel can be conceptualized more clearly.

References

1. Alcantara, E., Artacho, M.A., Gonzalez, J.C., Garcia, A.C.: Application of Product Semantics to Footwear Design. Part I - Identification of Footwear Semantic Space Applying Differential Semantics. International Journal of Industrial Ergonomics 35(8), 713–725 (2005)
2. Chuang, M.C., Ma, Y.C.: Expressing the Expected Product Images in Product Design of Micro-electronic Products. International Journal of Industrial Ergonomics 27(4), 233–245 (2001)
3. Demirbilek, O., Sener, B.: Product Design, Semantics and Emotional Response. Ergonomics 46(13/14), 1346–1360 (2003)
4. Griffin, A., Hauser, J.R.: The Voice of the Customer. Marketing Science 12(1), 1–27 (1992)
5. Helander, M.G., Khalid, H.M.: Affective and Pleasurable Design. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 3rd edn. Wiley Interscience, New York (2005)
6. Helander, M.G., Tham, M.P.: Hedonomics-Affective Human Factors Design. Ergonomics 46(13/14), 1269–1272 (2003)
7. Houser, K.W., Tiller, D.K.: Measuring the Subjective Response to Interior Lighting: Paired Comparisons and Semantic Differential Scaling. Lighting Research and Technology 35(3), 183–195 (2003)
8. Hsu, S.H., Chuang, M.C., Chang, C.C.: Semantic Differential Study of Designers’ and Users’ Product form Perception. International Journal of Industrial Ergonomics 25(4), 375–391 (2000)
9. Jindo, T., Hirasago, K.: Application Studies to Car Interior of Kansei Engineering. International Journal of Industrial Ergonomics 19, 105–114 (1997)
10. LaChance-Porter, S.: Impact of User Focus Groups on the Design of New Products. In: Proceedings of the 14th National on-line Meeting, New York, pp. 265–271 (1993)
11. Llinares, C., Page, A.: Application of Product Differential Semantics to Quantify Purchaser Perceptions in Housing Assessment. Building and Environment 42(7), 2488–2497 (2007)
12. Mondraqon, S., Company, P., Vergara, M.: Semantic Differential Applied to the Evaluation of Machine Tool Design. International Journal of Industrial Ergonomics 35(11), 1021–1029 (2005)
13. Nagamachi, M.: Introduction of Kansei Engineering. Japan Standard Association, Tokyo (1996)
14. Nagamachi, M.: Kansei Engineering in Consumer Product Design. Ergonomics in Design 10(2), 5–9 (2002)
15. Nakada, K.: Kansei Engineering Research on the Design of Construction Machinery. International Journal of Industrial Ergonomics 19, 129–146 (1997)

16. Osgood, C., Suci, G., Tannenbaum, P.: The Measurement of Meaning. University of Illinois Press, Urbana Champaign (1967)
17. Ou, L., Luo, M.R., Woodcock, A., Wright, A.: A Study of Colour Emotion and Colour Preference. Part 1: Colour emotions for single colours. Color Research and Application 29(3), 232–240 (2004)
18. Petiot, J., Yannou, B.: Measuring Consumer Perceptions for a Better Comprehension, Specification and Assessment of Product Semantics. International Journal of Industrial Ergonomics 33(6), 507–525 (2004)
19. Schütte, S., Eklund, J.: Design of Rocker Switches for Work-Vehicles: An Applications of Kansei Engineering. Applied Ergonomics 36, 557–567 (2005)
20. You, H., Ryu, T., Oh, K., Yun, M.H., Kim, K.J.: Development of Customer Satisfaction Models for Automotive Interior Materials. International Journal of Industrial Ergonomics 36(4), 323–330 (2006)
21. Yun, M.H., Han, S.H., Ryu, T., Yoo, K.: Determination of Critical Design Variables Based on the Characteristics of Product Image/Impression: Case Study of Office Chair Design. In: Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, pp. 712–716 (2001)
22. Yun, M.H., Han, S.H., Hong, S.W., Kim, J.: Incorporating User Satisfaction into the Look-and-feel of Mobile Phone Design. Ergonomics 46, 1423–1440 (2003)

Working toward Women's Economic Empowerment: Using Information and Communication Technology in Developing Areas to Market Traditional Crafts

Melissa Secore Levis

Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180 USA
melissa.secore@gmail.com

Abstract. Women face many challenges in regard to empowerment, especially economic empowerment. The United Nations (UN) has made a point in the Millennium Development Goals (MDGs) that work still needs to be done so women can reach some of the empowerment that their male counterparts enjoy. Traditional handicrafts are a rich part of indigenous women's heritage and these women should be enabled to market these items to a worldwide audience of customers. Information and Communication Technologies (ICTs) can be used to help women accomplish this. With proper dedication, education, and training, women can reach a new level of economic empowerment, along with other benefits, by selling their items via the Internet and other ICT-based technologies.

Keywords: Human-Computer Interaction, Crafts, women's empowerment, indigenous women.

1 Introduction

While women's economic empowerment has long been researched and worked on by governments and Non-Governmental Organizations (NGOs) worldwide, women still have a long way to come to realize the same level of economic empowerment as males. In particular, women face numerous barriers and obstacles when it comes to setting up and running businesses. The United Nations (UN) issued a list of Millennium Development Goals (MGDs) in 2000 that strive to reach gender equality by the year 2015[1]. The UN and its members have made women's empowerment a direct objective and ICTs can be used to help women achieve some of the empowerment many of the world's leaders believe they deserve.

Many women produce crafts that have been passed down through generations and can help them build "cottage industries." These can help bring in additional income to their families, and in some cases even fully support them. One way in which women can gain additional family income and create more opportunities for themselves is to market their culture's indigenous crafts using the Internet. Furthermore, women can also use ICTs (such as mobile technology) to help purchase supplies, find markets, and gain information. This paper will first provide a review of ICT projects that

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 266–273, 2011.

present case studies on the successes and failures of previous ICT-based craft projects. To also help better define the situation, issues regarding crafting, ICT use, and women's issues will be defined. The conclusion will include suggestions to better help women advance their economic empowerment though ICTs.

2 ICT Projects

There have been many craft and ICT-based projects launched around the world by a variety of organizations to help marginalized women. Such projects may provide job training, web design skills, access to education, and opportunities to obtain microcredit. While some examples do not directly relate to craft marketing and sales, it is important to examine these projects since they help women build their businesses and gain empowerment.

Women in many developing areas have been afforded new opportunities for economic empowerment because of ICTs. A very successful project was developed by the Grameen Bank which was founded in the 1970's in Bangladesh, a very poor, densely populated and underdeveloped nation [2]. Under its founder, Muhammad Yunus the bank pioneered the idea of microcredit and developed multiple initiatives that directly helped women. In 1997, Grameen Telecom was founded which led to the "village cell phone lady" phenomenon. As developing areas lack the infrastructure of land phone lines, mobile technology has achieved higher penetration [3]. The bank would issue credit to a woman who would take a loan, purchase a wireless phone, and buy minutes at wholesale prices, which would then turn around and re-sell to her community for a profit. This model of owning and using technology benefitted communities in multiple ways. The "cell phone ladies" were soon making almost three times the national average income in Bangladesh, therefore better providing for their families [2].

Maier and Nair-Reichert (2007) point out that ICTs are a "key solution for comprehensive development, poverty eradication and the empowerment of historically disadvantaged groups" [4]. The article also highlights the Grameen village phone project along with other successful ICT initiatives, such as projects that teach women computer skills, including desktop publishing at a computer microenterprise in Kasargod, Kerala, India [4]. This project provided women with IT skills, business skills, and networking opportunities and raised their confidence to better compete in the economy [4]. The authors highlight nine other successful ICT-based women's initiatives and explain how they succeeded. It should be noted that the article reviews one project, a weaving collective in Guyana that was successful at its peak, but has been struggling in more recent times due to a "coup" by male regional leaders who felt threatened and took over the project [4].

3 Women Crafting in Developing Areas

Crafting has long been a part of culture internationally and there are many stories of women participating in craft production (usually in collectives) to make money. These projects may or may not use ICTs to market their goods. Examples of women who have broken through traditional social norms and gained income range from an independent, unmarried woman named Mami in India getting funding to run a small craft-based company [5] to large weaving collectives in Morocco. Women have realized some great successes while making crafts, but there are also many examples of women who have faced enormous challenges. Some projects get taken over by men who feel threatened by women's successes [4], lack funding, or face an insufficient time investment, or the participants lack appropriate business training.

It is important to note that one reason crafting has empowered women is that women are able to participate in craft production activities during periods of discontinuous time; women can create these often dismissed "kitchen crafts" when it suits their schedule. This flexibility is important since in many cultures women are responsible for household chores and caring for their children and/or elders.

Women's craft production is very important not just from an economic standpoint, but also from a traditional, historical, and societal perspective. It is unfortunate that some younger women are not always interested in participating in craft traditions today [6]; this phenomenon puts these traditions at risk of being lost. An online archive of craft items may be a way for cultures to remember their rich crafting history if the production of these artifacts wanes. Joseph and Narayanan (2001) suggest that crafts being prepared for the online marketplace need are cataloged [7]. This activity creates a "digital catalogue" for a culture's indigenous craft items [8].

Crafting often gives information or insight into a culture; the artifacts themselves can tell stories. For example, during the Pinochet dictatorship in Chile, some women made what are called *arpilleras*. These three dimensional quilt-like wall hangings were made by women to express their frustrations with the political situation, often in terms of remembering a missing family member. A common message displayed on these pieces was ¿Dónde están?" (Where are they?) [9]. These crafts were dismissed by many as being mere "women's work" even though many of them carried a political message [9]. During the Pinochet dictatorship, women made *arpilleras* in secret. Today, the political climate in Chile is different. Women now sign the backs of their *arpilleras* with their names and often "AFDD" *Agrupación de los Familiares de los Detenidos Desaparecidos* (Association of the Relatives of the Disappeared) [10]. Today many *arpilleras* even depict non-political scenes. Examples of the more "modern" *arpilleras* are easily found for sale on the Internet and they can fetch a substantial price. There are certainly opportunities for women to market *arpilleras* online since these wall hangings are often popular in children's room décor [10].

Nonetheless, women face a variety of challenges and barriers regarding craft production and marketing. One hurdle that women face in some areas is that certain crafting activities are reserved for men only. There are cases of men being the only ones who are able to receive training in craft production. For instance, in Orissa, India, only men and blood-related women are able to receive certain types craft training from gurus [5].

Middlemen also pose a tremendous challenge to women crafters. In non-ICT-based craft trade, middlemen are often able to take advantage of women crafters. According to Agnes Pala, a west Kenyan woman now residing in New York State, middlemen go into rural areas to purchase handmade crafts when women are most in need of money, such as when school fees are due. Since women are most vulnerable during these times, they may not even make back their cost on an item, let alone make any money for their time and skills. As described in Goldstein and O'Connor's article (2000), middlemen have long enjoyed an economic advantage. But women can use ICTs to gain an advantage over middlemen by having access to more accurate pricing. In doing this, women can leverage better pricing and even eliminate some or all of the "layers" of middlemen [11].

4 Women Using ICTs in Developing Areas

Giving women the tools to take control of both the production and marketing of crafts is critical to their economic empowerment. As Yunus (1998) said regarding people and ICTs, "They must be owners of that technology, not just its passive consumers" [2]. Many developing nations continue to face major technical barriers to internet access for all or some groups of their populations. Women in developing areas may deal with another major obstacle when it comes to ICT access: lack of access due to gender. ICT access can be a much larger challenge for women than men as many social and cultural factors limit women from using the Internet and other communication technologies. For example, in some cultures women feel uncomfortable in public areas, which restrict them from using the internet in public areas such as cybercafés [12]. Also, women do not have much time to learn and use ICTs. Women are usually the ones who provide domestic care for children and other family members and these responsibilities often leave them with little extra time to use ICTs to build a business or gain education. When women do have some "spare" time, that they may be able to use for ICTs, it tends to be in the evening when traveling to and from cybercafés or kiosks may not be safe [12]. There is also a lack of training and education for women in these tools. Just as in Orissa India, where some women are kept from crafting, many women in the world are kept from ICTs [5].

Many organizations and even governments are interested in changing the situation. As part of its MDGs the UN has raised awareness and to attempted to build policies around the importance of ICTs in gender equality. The UN strongly advocates women use a variety of ICT-based devices/modes to improve their lives and increase employment opportunities [13]. The more traditional forms of ICT, such as radio and television may serve as successful one-way communication methods, helping women receive educational content in their homes. However, for women to truly become producers and earn income, they need to both to receive and disseminate information. To fully participate in commerce, they have to retrieve information (e.g. get information about pricing) and send information (e.g. put goods online for sale.) If women can use the Internet to accomplish these activities, they can expand their market though e-commerce. Increasingly, especially in African countries, mobile phones are becoming a valuable tool and supplementing the internet for commerce, which is referred to as m-commerce [13]. In many areas mobile cell phone technology is more available to people than computers. In 2007, 10 per cent of South Africans had access to the Internet via computers, versus the 87 per cent who had mobile phone service subscriptions [3].

Mobile phones may not be a direct marketplace for consumers to buy finished craft goods, but they can be extremely powerful when used as a business tool by women producing craft items. Mobile phones let women communicate with other business people and get up-to-date information on fair pricing. On the grassroots level, mobile technology is also being used to disseminate information to women, especially those with less access to information than males [3]. Both e- and m-commerce can greatly increase women's economic growth and empowerment; they both allow women to have access to training and market information, save on travel costs, get access to loans, and improve communication with customers [13].

Not only can ICTs help women sell their goods, but the education regarding ICTs can give women the valuable skills they need to use the Internet and other media accomplish this. Women need education on ICTs and business skills. Distance education, sometimes referred to as e-learning or e-education, is a way for busy women to gain education. Some focus has already been placed on academic education since there is a wide gap in many developing areas. As of 2002, only 3 percent of students in Sub-Saharan Africa and only 7 percent in Asia received some form of postsecondary education [14]. Less-traditional students (e.g. adults) may benefit more from educational opportunities delivered via ICTs. For instance, in Brazil the Committee to Democratise Information Technology (CDI) created over 100 schools that teach students ICT skills. Remarkably, over 25,000 students per year are trained by these self-sustaining schools that depend on volunteers, limited funds and used computer equipment [14].

Indigenous women crafters can also use ICTs to improve their craft techniques [15]. Resources available via ICTs can offer new insights, give advice for greater efficiency, and even offer training and instructions for making new crafts. Craft traditions should not be divorced from modern life. New craft artifacts can be created if women use ICTs to discover new and innovative processes and ideas.

5 Bringing It All Together: Women Using ICTs to Market Crafts

Women face issues of funding if they want to start their own businesses, especially businesses that depend on expensive ICTs. Accessing the Internet and other ICTs is expensive and may be cost prohibitive in developing areas. Funding and credit are essential factors to help women create these successful opportunities for themselves. In many cases, women can use mobile ICTs to gain better access to funding and credit sources.

Microcredit and microloans have helped some create successful business opportunities over the past several years. The Grameen model has been implemented in different developing areas. There are also other cases of microcredit helping women run micro, small and medium enterprises in developing areas. Both governments and NGOs need to be involved in helping women crafters apply for, obtain, and resourcefully use the funds they receive. A challenge that women face even after they successfully obtain loan money is that after their businesses take off and become larger, male family members may take over the businesses [6]. There needs to be careful follow up and continuing education and training to ensure that the women who work hard to build businesses can continue to run them and enjoy their benefits. For example, in Malaysia women make up a smaller percentage of the workforce and an even smaller percentage of managers and employers. In the year 2000, only 10.4 percent of Malaysia's female population were employers [6]. Due to this disparity, many government organizations have taken deliberate steps to create business opportunities for women. In Malaysia grants, credit, and other support services have been offered to women for over two decades. Unfortunately, the grant programs showed low disbursement rates. According to Shah (2008), this was due to a low number of applications. Women were often overwhelmed by the amount of work that was necessary to secure the grant funds [6].

Credit is slowly getting easier for women to obtain, but it is still a challenge in many cases. Women can gain more access to microcredit and microloans if they band together and form collectives. Furthermore, as Pala (2010) mentioned, financial institutions need to keep moving in the direction of taking alternate forms of collateral, such as bicycles and rickshaws.

Online, many existing infrastructures are already globally available to indigenous women crafters to allow them to sell their merchandise. ICTs can also help women crafters eliminate middlemen from their economic cycle and take full control of the income. Women can take advantage of the structures already set up on the Internet to market their crafts, such as auction websites and collaborative artist websites. However, it should be noted that these sites are not perfect and do have their issues. Sites such as www.tenthousandvillages.com, www.aidtoartisans.org, and www. novica.com offer opportunities for crafters to sell their goods. Furthermore, Aid to Artisans has a growing research agenda and offers business training, product development and market access to crafters. These preexisting sites and organizations already help women market crafts. It's not always necessary to reinvent the wheel, and that is certainly true in this case. Web development takes time and specialized skills. Indigenous women do not need to develop programming skills; having pre-set templates and websites will allow women to focus more on craft producing activities and business development.

It needs to be noted that it can be difficult for women to start using preexisting e-commerce sites, especially auctions. Women who are new users do not have the feedback built up and risk seeming untrustworthy to customers. Also, many western customers have specific quality standards. Pala (2010) said that she and other merchants using EBay for auctions and PayPal needed to have a large amount of money stored in their accounts to be able to supply rapid refunds to unsatisfied customers. Such a cash reserve is a determent to participation by most indigenous women. It is also important that international women crafters have access to trustworthy channels that offer secure financial transactions. Finally, women in developing areas may not have sufficient ICTs to regularly monitor results.

Consider a fictional woman named Esther living in Kenya. She has extensive experience weaving baskets that western people like to use as purses. She wants to create a better economic situation for herself and her children by selling her handmade baskets, but middlemen have been taking advantage of her and have not been offering her fair prices. Esther owns a cellphone and has access a couple of hours a week to a kiosk computer. She has no money of her own to invest into her business, but does own a bicycle. What can she do to make a fair price on her goods? First, she'll need some funding. She reaches out to a microlending organization that her cousin told her about. She does this by obtaining the organization's phone number online, then reaching out to them on her cellphone. Esther reaches an agreement with the lender by putting her bicycle up as collateral for a small loan that will allow her to buy some basket making supplies and for items and services she'll need later. She has a delay before she has to start making payment on her loan and she gets to work making a supply of baskets. The lending organization sends her weekly text messages reminding her when her first payment is due.

When it comes time to sell her baskets, the middleman is out of the question. He is not even willing to pay her what she spent to make a basket. Fortunately, the lending organization she's affiliated with put her in touch with a women's crafting collective that sells African craft items on EBay. The organization loans her a digital camera (which she makes a small, refundable deposit on from her loan money) to take pictures of her goods. Esther then uses the kiosk computer to send her photos to the craft collective. A volunteer with the collective coaches Esther over her cell phone though the upload process. Esther also verbally dictates item descriptions to the volunteer. The collective then posts the images and item descriptions on EBay. When one of Esther's items sells, it is her responsibility to send it out to the customer. The customers pay shipping and handling. Once the collective receives the item payment into their PayPal account and Esther confirms she has sent the item, the collective transfers Esther's money into her personal PayPal account, minus a modest administration fee that helps fund the collective. A critical part to all this is that the collective has money set aside in case a customer requests a refund. In a short time, Esther is able make payments on her loan, plus interest.

This scenario demonstrates how women do not need to be technology experts to use ICTs to make money from their crafts. Pointing and shooting a camera, using a cell phone, and learning basic computer skills are all a women needs in addition to minimal literacy and her traditional craft skills. Modest startup funding from governments and NGOs can help make fictional scenarios a reality.

While these suggestions open many doors for women crafters, there are still overhanging challenges. Language, literacy, and context issues remain in play. But with the increasing number of organizations that are offering partnerships these barriers can be overcome. ICTs are making the world seem smaller, and women can receive remote assistance from a global pool of volunteers. These volunteers may offer technical education and support and link women with similar interests together [16].

6 Conclusion

Women's crafting makes up a small part of women's economic growth and empowerment. However it has seen successes for the most vulnerable and is an area that continues to demand attention by governments and NGOs alike. Excitingly, Pala (2010) mentioned that these opportunities can change a woman and family's life. Women producing "kitchen crafts" that they once sold to middlemen can develop viable and profitable small and medium sized businesses. Governments, lenders, and NGOs that provide business skill education and ICT tools help women take control of their finances, creating better opportunities for themselves and their families. While there have been some great successes, there is still much work to be done to put more power into the hands of women and to meet the UN's MDGs. Better understanding of the constraints on women's use of ICTs can help.

References

1. United Nations Millennium Development Goals, `http://www.un.org/millenniumgoals/`
2. Yunus, M.: Alleviating poverty through technology. Science 282, 409–410 (1998)
3. Joseph, M., Andrew, T.: Information and communication technology policy imperatives for rural women empowerment: focus on South Africa. In: IEEE Africon, pp. 1–6. IEEE Press, New York (2009)
4. Maier, S., Nair-Reichert, U.: Empowering women through ICT-based business initiatives: an overview of best practices in e-commerce/e-retailing projects. Information Technologies and International Development 4, 43–60 (2007)
5. Acharya, J., Lund, R.: Gendered spaces – socio-spatial relations of self-employed women in craft production, Orissa, India. Norwegian Journal of Geography 56, 207–218 (2002)
6. Joseph, L., Narayanan.: `http://www.foodindia.org.in/emarketers/narrative.htm`
7. Shah, Farida.: Mainstreaming Potential Women Exporters in International Markets through ICT: Malaysia. In: APEC Committee on Trade and Investment (2004)
8. Ssewanyana, J.K.: ICT Access and Poverty in Uganda. International Journal of Computing and ICT Research 1(2), 10–19 (2008)
9. The Politics of Chilean Arpilleras, `http://www.thejunctionni.org/Chilean%20Catalogue%20Complete.htm`
10. Chilean Arpilleras: A Chapter of History Written on Cloth, `http://cachandochile.wordpress.com/2010/09/11/chilean-arpilleras-a-chap`
11. E-commerce for Development, `http://www.oecd.org/dataoecd/37/61/1922730.pdf`
12. Hafkin, N.: Gender Issues in ICT Policy in Developing Countries: An Overview. Division for the Advancement of Women (DAW) Expert Group Meeting on Information and communication technologies and their impact on and use as an instrument for the advancement and empowerment of women (2002)
13. Note by the Secretariat, Information and Communications Technology and Gender Equality: New Opportunities and Challenges for Public Administration to Implement the Internationally Agreed Development Goals, Including the Millennium Development Goals. Public Administration Perspective on Implementing the Internationally Agreed Goals and Commitments in Regard to Gender Equality and Empowerment of Women (2010)
14. A Framework for Poverty Alleviation with ICTs, `http://www.tanzaniagateway.org/docs/framework_for_poverty_alleviation_with_ICTs.pdf`
15. Bonder, G.: From Access to Appropriation: Women and ICT policies in Latin American and the Caribbean. Symposium conducted at the meeting of the Division for the Advancement of Women (DAW) Expert Group Meeting on Information and Communication Technologies and their Impact on and use as an Instrument for the Advancement and Empowerment of Women (2002)
16. Pattanaik, D.: Engendering Knowledge Networks – Empowering Women through ICT. In: ACM Women and ICT (2006)

Socio Economic Psycho Knowledge Based Intelligent Agents for Automated e-Commerce Negotiation

P. Vijayaraghavan[1,2] and R.Ponnusamy[1]

[1] Department of Computer Science & Engineering
Aarupadai Veedu Institute of Technology, Chennai, Tamilnadu, India
[2] Department of Computer Science & Engineering
Jawarhalal Nehru Technological University, Kakinada, Andhrapradesh, India
vijayaragavanpm@avit.ac.in, rponnusamy@acm.org

Abstract. Automated Negotiation is a process in which two or more parties with different criteria, constraints, and preferences, jointly reach an agreement on the terms of a transaction through an automated constraint satisfaction and preference selection mechanism. In practical real environment the process of negotiation is not very strictly a mechanical selection or constraint satisfaction. It is based socio economic psycho conditions. One-to-many negotiation framework is taken as the default to test this socio economic psycho knowledge incorporation. In these cases of auctions, online trading needs a greater flexibility. Therefore it needs different strategies with different opponents. In this paper, an attempt is made to incorporate the socio economic psycho knowledge in the process of negotiation. All agents working on behalf of one party negotiate individually with other parties. After each negotiation cycle, the group of these agents report back to a coordinating agent that evaluates how well each agent has done, and issues new instructions accordingly. Each individual agent conducts reasoning by using socio psycho knowledge constraint-based technique with the objective of overall profit for both sides. We outline two levels of strategies that can be exercised on two levels, the individual negotiation level, and the coordination level. In our prototype Intelligent Trading Agency (ITA), agents autonomously negotiate multi-attribute with socio economic psycho consideration terms of transactions in an e-commerce environment tested with a personal computer trading scenario.

Keywords: Automated Negotiation, Intelligent Trading Agency, Socio-economic Psycho Knowledge.

1 Introduction

Electronic Commerce [1] has unleashed yet another revolution, which is changing the way businesses buy and sell products and services. Modern age of busy life leaves no time space for negotiated purchase. A business organization can organize itself to conduct e-commerce with its customers. The Business-to-Customer or Customer-to-Customer trading technique requires different automated tools and mechanisms to assist the customer over the trading environment to save time and achieve their

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 274–284, 2011.

objectives in understanding practical conditions. As a result we try to develop new methodologies for automated decision making and negotiation strategies.

Online trading provides various types of sophisticated facilities for different products procurement and sales. The application of multi-agent [2,3,4] in automatic negotiation focus on negotiation process design, environment and its three dimensions, negotiation protocol, model and strategy under the bilateral negotiation with multi-attribute. These software agents can also play an important role in providing automation and support for the negotiation stage of online trading [5]. Different types of negotiation scenarios exist ranging from one-to-one to one-to-many, many-to-many [6]. In this case only very limited knowledge is used for negotiation and most of the approaches consider only three or four attributes for negotiation.

Cyber psychology [7, 8] explores the psychological aspects of environments created by computers and online networks. It presents an evolving conceptual framework for understanding how people react to and behave within cyberspace: what is called as "the psychology of cyberspace" - or simply "cyber psychology”. Most existing e-commerce system’s negotiation mechanisms have been focusing on the technical efficiency and adopting mechanical strategies. It fails to adapt the practical conditions that exist in the real time environment. In the real time environment negotiation is based on the socio economic conditions. This socio economic condition adaptation enables the people to become successful in their business environment. An even person thinks about the win-win situation of adopting the condition overall profitability and long term business strategy and other customer relation maintenance are considered. Similarly in the present design an attempt made to simulate the automated negotiation strategy in adopting the socio economic cyber psycho knowledge business conditions for negotiation.

In this paper an attempt has been made to adapt all these socio-economic psycho knowledge for negotiation. As per the negotiation scenario model is concerned it uses a multi-attribute one-to-many negotiation strategy. Also, it already proved that the face-to-face negotiation [9] was not different than e-negotiation, in terms of the final price. However, both the negotiation media and the negotiation sequence significantly affected the main features of the negotiation process, in terms of time duration and the use of hard or soft tactics. Incorporation of cyber psycho enables us to adapt these tactics.

This paper is organized as follows. The section 2 explains the negotiation as a distributed constraint satisfaction problem. Section 3 explains the one-to-many negotiation mechanism. Section 4 explains the impact of socio economic psycho knowledge. Section 5 gives simple prototype implementation of one-to-many negotiation for purchasing a camcorder. Section 6 concludes this paper.

2 Negotiation as Distributed Constraint Satisfaction with the Adaptation of Socio Economic Psycho Knowledge

Negotiation is a dialogue intended to resolve disputes, to produce an agreement upon courses of action, to bargain for individual or collective advantage, or to craft outcomes to satisfy various interests. It is the primary method of alternative dispute resolution to reach a consensus. Effective negotiation helps you to resolve situations where what you want is in conflict with what someone else wants. The aim of win-win negotiation is to find a solution that is acceptable to both parties, and leaves both parties feeling that they've won, in some way, after the event. In general, negotiation can be classified according to the number of parties involved and the number of attributes negotiated. In terms of the parties involved, negotiation scenarios can be one-to-one, one-to-many or many-to-many. In terms of negotiation attributes, a negotiation can involve a single attribute (eg. Prize) or multiple attributes (eg. Prize, quality and delivery time). In this attempt intelligent trading Agency supports multi-attribute, one-to-many negotiation and incorporate the cyber psycho-economic knowledge.

2.1 Cyber Psychos-Economic Knowledge Based Seller and Buyer Agents

Incorporation of cyber psycho economic knowledge is through the buyer and seller agent. As per the seller is concerned it must collect the following knowledge about the buyer agent. This information are buyer group information, customer economic status, product margin, product movement index and overall profitability. Virtual communities [10] represent an effective way for facilitating the circulation of knowledge in organization and groups. Particularly, this approach relies on using cognitive agent informed of social cognition theories that are able 1. to infer the individual participatory profile of the members from the observation of their online behavior 2. To use these profiles and the participatory principles to determine individual interventions that are the more likely to impact people participation 3. To intervene proactively.

1. Buyer (Customer) Group information. As like above, in this case we classify the different types of customers (Buying Agents). Such as Regular customer, new customer based on their history. If customer agent is regular then Ti is assumed to be 1, otherwise it is 0.5.

2. Customer economic status. In another method, different economic states are identified. Based on this, the Standard Customer, Non-standard Customer and Bulk Customer are identified. The Standard customer is that customer who will not bargain much on prize. They will bargain only for other attributes. The bulk customer is a customer who buys a huge set of products in one or two instances. These two above attributes are considered for negotiation. If customer agent is standard customer/bulk customer then Ei is assumed to be 1, otherwise it is 0.5.

3. Margin. This attributes indicates the maximum margin prize that can be offered to the customer. It is also one of the important attributes considered for negotiation. If seller margin is above 20% then Mi is assumed to be 1, otherwise it is 0.5.

4. Product movement index and future sales ratio. An expert marketing professional maintain a Product Movement Index (PMI). This indicates the speedy movement of the product in the market. If the market is good then the index will be high. Some time the high discount offer is given to the low Product Movement Index Products. If PMI is good then PMIi is assumed to be 1, otherwise it is 0.5.

5. Over all profitability. Over all products set profitability is taken for the complete set of the product. Sometime the high offer can be given to the last few products without loosing the overall profitability of the product set. If the overall profitability is good then Pi =1, otherwise it is assumed to be 0.5.

Table 1. Customer Profile

Customer Id	**#00078**
Regular Customer	Yes
Standard Customer	No

Table 2. Product Profile

Product Id	**#0089**
Product Margin	30%
PMI	High
Overall Profitability	High

2.2 Offer Evaluation

While considering these additional attributes of socio-economic psycho knowledge, the customer will be treated differently. They will be given different offers based upon the socio-economic psycho status. Even though these attributes are considered, the main objective is to increase overall profitability of both sides. ITA agents use multi-attribute utility theory and constraint-based reasoning [11] for the evaluation and generation of offers and counteroffers. Let us start with offer evaluation. For an offer received by agent A from agent B to be considered, it has to satisfy all constraints, that is, the proposed value of each variable must belong to its domain as specified by agent A. Then, the value of an offer consisting of a number of attributes X = {x1,x2,……xn} is defined as a function.

$$v(x_1,x_2,\ldots\ldots x_n) = \sum_{i=1\ldots n} w_i v_i(x_i) / \sum_{i=1\ldots n} w_i = c_i \qquad (1)$$

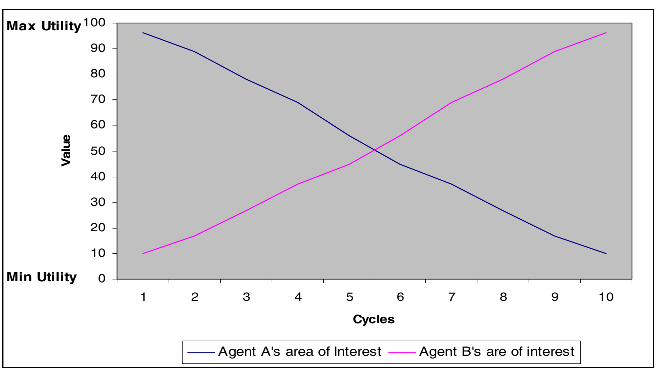


Fig. 1. Constraint-Based Model of One-to-One Negotiation

Where xi is ith attribute of negotiation, vi is the utility function of the ith attribute and w_i is the weight (priority) of the value of the ith attribute. c_i is the socio-economic psycho weight of the ith customer's. The utility function is used for comparing and ordering alternative acceptable solutions.

Fig. 1 shows that the acceptable offer by both agents is one which satisfies at least the minimum utilities of both agents. If the line meets at a particular point, a solution can be found, otherwise the negotiation fails. An agent can see the minimum boundary as a constraint on the domain of acceptable total utility values.

2.3 Offer Generation

In this case of ITA both agents generate the offers and counter offers by adapting different strategies. Different agent negotiation strategies are as follows

i. If the customer is the standard customer no offer is given to him, otherwise if he is the non-standard customer then the offer is given to him in a step by step. If the customer is the bulk customer then the offer is specially given to him.
ii. If the customer is the regular customer then the offer is given to him otherwise no special offer is given to him.
iii. If the product movement index is good and the margin is good then the offers is given to him, otherwise no offer is given.
iv. The overall profit margin is good then the offer is given, Otherwise no offer is given to the customer.

The overall socio-economic psycho weight

$$c_{i\,(1,2,\ldots\ldots n)} = 1 - T_i E_i M_i P_i PMI_i (y) \tag{2}$$

Table 3. Customer Personality Type Classification

S. No.	Personality Type	Basic Attitude	Ti
1.	Psychopathic (antisocial)	Don't want to purchase anything	0.1
2.	Depressive and manic (impulsive)	Unpredictable Customer with less possibility	0.2
3.	Hysterical (histrionic)	Unpredictable Customer reasonable possibility	0.3
4.	Dissociative	Unpredictable Customer high possibility	0.4
5.	Masochistic (self-defeating)	Based on the attractive Advertisement he will buy the product	0.5
6.	Schizoid	Want to purchase things with high profitability	0.6
7.	Narcissistic	Want to purchase things with reasonable profitability	0.7
8.	Paranoid	Somehow he want to purchase	0.8
9.	Obsessive and compulsive	Somehow without considering the cost he/she will buy the product	1.0

The socio-economic psycho weight,c_i for product y at i^{th} offer is the combined parameter of Ti Ei Mi Pi PMIi, where Ti indicates the user type, Ei indicates the economic status, Mi indicates the margin, Pi indicates the overall profitability and PMIi is the product movement index. The customers are classified in to different types based on their personality type and is given in the following Table 3. So, the offer is given based on all these combined parameters.

If agents want to lead effectively, agent needs to be able to make good decisions. If agents can learn to do this in a timely and well-considered way, then the agent can lead best team to spectacular and well-deserved success. However, if agent dither or make poor decisions, your team risks failure and agent time as a leader will probably be brutally short. The techniques in this section help the agent to make the best decisions possible with the available information agents have. They help agents map out the likely consequences of decisions, work out the importance of individual factors, and choose the best courses of action.

3 One-to-Many Negotiation Mechanism

In this present approach to automate one-to-many negotiations, the Iyad Rahwan's [6] approach is adapted in order to test the impact socio-economic psycho knowledge incorporation. The main advantage of this approach it is possible to scale to many-to-may negotiation.

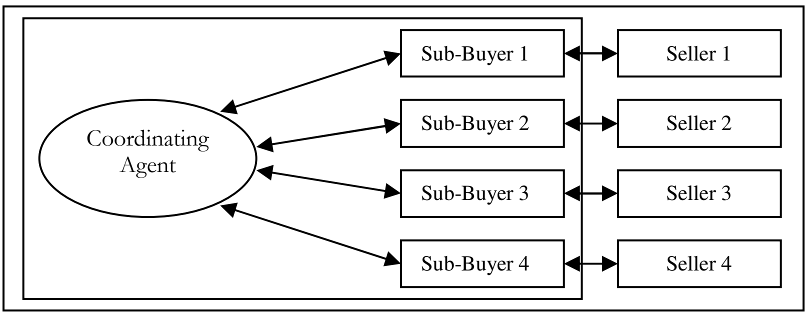


Fig. 2. One-to-Many negotiation (One buyer & many sellers)

We now present our approach to automating one-to-many negotiations. The scenario we are tackling is one in which one agent (buyer or seller) wants to negotiate a deal with a number of opponents, in order to find the best possible deal in the market. We propose reusing the techniques and components we used in one-to-one negotiations. This offers an advantage over approaches in which one single complex agent must conduct and directly maintain multiple threads of negotiation. In ITA, an agent can negotiate with many other agents by creating a number of one-to-one negotiating agents that negotiate on its behalf, and perform the task of coordinating them. We will call these agents sub-negotiators. Every sub-negotiator conducts a one-to-one negotiation with a different opponent. After each negotiation cycle (one offer and counteroffer), each sub-negotiator reports the results back to the coordinating agent. The coordinating agent then evaluates the situation, and issues instructions accordingly.

Fig. 2 shows an instance of one-to-many negotiation scenarios. In this particular scenario, a buyer agent negotiates a deal with many prospective sellers. The buyer agent consists of a coordinating agent and a number of sub-negotiators (sub-buyers). All sub-buyers represent the preferences and constraints of the same buyer, but they may use different negotiation strategies. Similarly, a single selling agent can negotiate with a number of prospective buyers by instantiating a number of sub-sellers, and coordinating them. From an architectural point of view, our approach has many advantages over existing systems:

i. It offers simplicity and reusability by allowing us to reuse any existing one-to-one negotiating agent in a one-to-many setting, hence providing rapid development of negotiation applications.
ii. It allows for the system to be highly customizable since sub-negotiating agents can be modified, removed, or new agents with new strategies and capabilities can be added dynamically to the system at any point in time.
iii. This also allows for better scalability since not only can the different negotiation parties be on different machines over a network, but also can the different sub-negotiators of the same agent.
iv. The resulting system becomes more robust compared to a centralized complex agent. If one negotiation thread dies due to technical difficulties for example, the other threads can continue (as long as the coordinating agent is still alive).
v. In principle, it would be also possible for each sub-negotiator to be a one-to-many negotiating agent consisting of a coordinator and several sub-negotiators, and so on.

In ITA, there are two levels of negotiation strategies, namely strategies exercised by individual buyer or seller agents in their one-to-one encounters, and strategies exercised by the coordinating agents in organizing and issuing commands to their sub-negotiators. Negotiation Strategies of individual sub-negotiators include:

i. Take it or leave it (fixed offer).
ii. No concession (same level of satisfaction, but possibly different offers).
iii. Fixed concession.
iv. Better deal strategies (resuming negotiation after deal is found hoping a better deal may come).

Details of these strategies and a few experiments on them can be found in [12]. We outline a few simple coordination strategies that can be exercised by the coordinating agent for controlling sub-negotiators:

i. Desperate Strategy: This is a very simple strategy in which the time constraints may be important and the agent wants to close a deal fast. In this strategy, as soon as a sub-negotiator finds an acceptable offer, the coordinating agent accepts it and sends messages to all other sub-negotiators to terminate their negotiation. If more than one sub-negotiator comes up with an acceptable offer, the one with the highest utility is chosen while the rest are terminated.
ii. Patient Strategy: In this strategy, even if an acceptable deal is found by one or more sub-negotiator(s), those agents are asked to wait while all other agents are asked to resume their negotiations. Once all sub-negotiators complete their negotiation process (whether with success or failure), the best offer is chosen. This strategy guarantees that the best possible deal can be reached, but does not give regard to time constraints. This might be a significant limitation in a marketplace with too many potential suppliers to negotiate with. One variation of the patient

strategy is one in which a time limit is set by the user, within which if no better deal is found, the negotiation terminates and the best deal so far wins.

iii. Optimized Patient Strategy: In this strategy, the coordinating agent uses information about one negotiation outcome to influence the performance of other sub-negotiators. The constraint on the utility for the other sub-negotiators is updated in order to avoid unnecessary deals which are not as good as the one already found. For example, if the accepted minimum total utility is 5, and one sub-negotiator has found a deal with utility 7, there is no point in other sub-negotiators reporting back a deal with utility 6 even though it is an acceptable deal (according to the initial constraints). In this case, the constraint on the utility for all remaining sub- negotiators is updated to be 7, causing any deal below that margin to be unacceptable. This also ensures that no sub-negotiator offers an offer that is worse than an offer received by a fellow sub-negotiator.

iv. Strategy Manipulation Strategies: In this class of strategies, the coordinating agent may modify the negotiation strategies of different sub-negotiators at runtime. For example, after securing a deal, other sub-negotiators can exercise a take-it-or-leave-it strategy with their opponents. More sophisticated use of such strategies is left for future research. In addition to the architectural advantages mentioned above, there are advantages relating to the method of representing preferences and constraints. We are using declarative knowledge representation in the form of constraints, which can be easily exchanged between, and understood by, different agents, makes adding and removing sub-negotiators and communication between them an easier task.

4 Prototype Simulation

In this section, a prototype implementation to demonstrate ITA's capabilities are tested with Java based IBM aglets [13]. An Agent Transfer Protocol (ATP) is used to communicate with these different agents. Aglet uses a technique called serialization to transmit data on the heap and to migrate the interpretable byte-code. These aglets are supporting message passing and broadcasting. Each aglet is integrated with the functional components of this architecture. The blackboard system is shown as the explicit component and is implemented through using standard java serialization.

After execution the coordinating agent is able to automatically select the sub-buyer after comparing their resultant negotiation. In this case, all the plot screens are shown in Fig. 3,4,5,6. The program is underdevelopment and so the values are taken and screen is plotted through MS-Excel. From these four sub-buyers the agent is able to

Table 4. Socio-Economic Psycho Weight C_i Matrix for Sample Scenario

Sub-Agent No. (i)	Ti	Ei	Mi	Pi	PMIi
1	0.5	1	0.5	1	1
2	0.5	1	1	1	1
3	1	1	1	1	1
4	0.5	1	0.5	1	0.5

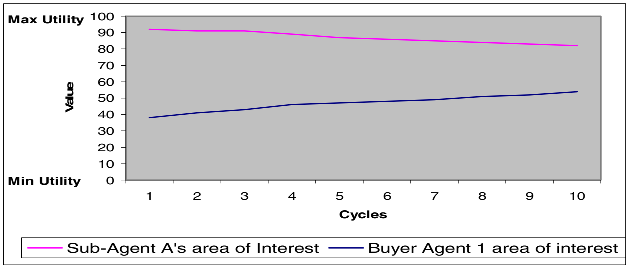


Fig. 3. Negotiation Result for Sub Agent A

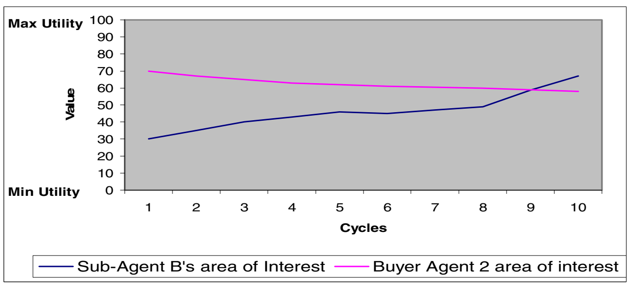


Fig. 4. Negotiation Result for Sub Agent B

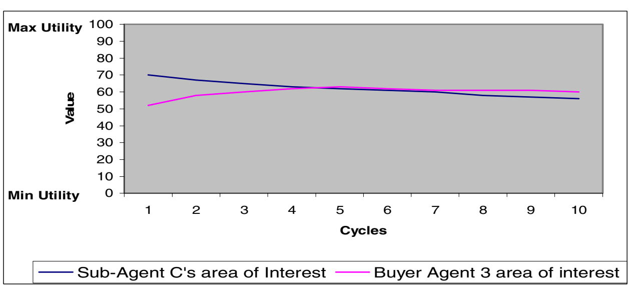


Fig. 5. Negotiation Result for Sub Agent C

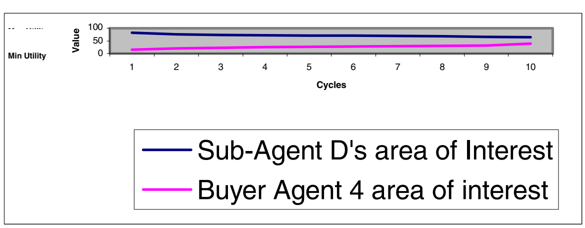


Fig. 6. Negotiation Result for Sub Agent D

select third sub-buyer. It gives the best cost profit after negotiation. Even though Sub-Agent B and Sub-Agent C are reaching a solution, Sub-Agent C is always chosen, because Sub-Agent C is having good Socio-Economic Psycho Weight.

5 Conclusion

This paper presents an intelligent trading agency (ITA) to support fully autonomous multi-attribute negotiations in the presence of socio economic psycho knowledge. It is shown that incorporation of socio economic psycho knowledge enables the system to give better profit for both the buyer and seller. The multi-agent based one-to-many negotiation is proved to be customer adaptive, business adaptive and automated. The results are compared with the regular constrains, preferences, criteria and socio economic psycho knowledge implementation. One individual agent will be able to negotiate multi-issue by exchanging offers and counter offers until they either reach a consensus that satisfies each party's private preferences and constraints, in adopting socio economic conditions, or they run out of offers and the negotiation fails. The system is implemented using the Aglet and a pilot application that uses a Camcorder trading scenario.

References

1. Bajaj, K.K., Nag, D.: E-Commerce: The Cutting Edge of Business, 2nd edn. Tata McGraw-Hill Publishing Company Ltd, New York (2008); ISBN-13:978-0-07-058556-0
2. Jennings, N.R., Faratin, P., Norman, T.J., O'Brien, P., Odgers, B., Alty, J.L.: Implementing a Business Process Management Sysem using ADEPT: A Real-World Case Study. International Journal of Applied Artificial Intelligence 14(5), 421–465 (2000)
3. Tang, X., Zhang, N., Gong, S.: Automatic Multi-Agent Negotiation in E-Procurement. In: International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2007, September 21-25, pp. 4686–4690 (2007)
4. Chavez, A., Maes, P.: An Agent Marketplace for buying and Selling Goods. In: Proceedings of the First International Conference on the Practical Application of Intelligent Agent and Multi-Agent Technology, London, UK (April 1996)
5. Guttman, R.H., Moukas, A., Maes, P.: Agents as Mediators in Electronic Commerce. Electronic Markets 8(1) (1998)
6. Rahwan, I., Kowalczyk, R., Pham, H.H.: Intelligent Agents for Automated One-to-Many e-commerce Negotiation. In: 25th Australian Computer Society Conferences, Melbourne, Australia (2002)
7. Suler, J.: The Psychology of Cyberspace (1996), `http://wwwusr.rider.edu/~suler/psycyber/psycyber.html`
8. Barak, A., Suler, J.: Reflections on the Psychology and Social Science of Cyberspace. In: Psychological Aspects of Cyberspace: Theory, Research, Applications, pp. 1–12. Cambridge University Press, Cambridge (2008)
9. Galin, A., Gross, M., Gosalker, G.: E-negotiation versus face-to-face negotiation what has changed – if anything? Computers in Human Behavior 23(1), 787–797 (2007)

10. Nabeth, T., Angehrn, A.A., Mittal, P.K., Roda, C.: Using Artificial Agents to Stimulate Participation in Virtual Communities. In: CLEDA, Portugal, December 14-16 (2005)
11. Yokoo, M., Durfee, E.H., Ishida, T., Kuwabara, K.: The Distributed Constraint Satisfaction Problem: Formalization and Algorithms. IEEE Transactions on Knowledge and Data Engineering 10(5) (September/October1998)
12. Kowalczyk, R., Bui, V.: On Constraint-based Reasoning in e-Negotiation Agents. In: Dignum, F., Cortés, U. (eds.) Agent Mediated Electronic Commerce III. LNCS (LNAI), pp. 31–46. Springer, Heidelberg (2000)
13. `http://www.trl.ibm.co.jp/aglets`

Shopping Cart Interactive Program (SCIP)

Cyndi Wiley, Emmanuel Saka, Stefan Tauber, and Sunghyun R. Kang

Iowa State University, Department of Art and Design,
146 College of Design, 50011 Iowa, USA
{clwiley,esaka,timetobu,shrkang}@iastate.edu

Abstract. The Shopping Cart Interactive Program (SCIP) application was designed to improve shoppers' connection with the grocery retail environment. Nutrition is central to human life. Eating the right foods can significantly impact one's quality of life. With the capabilities of the SCIP application to provide an interface that can display readable health information on a mobile device and tailor to the specific dietary restrictions of its users, it will enable shoppers to have a less frustrating experience when shopping for food.

Keywords: Interface design, User-centered design.

1 Introduction

Our population is significantly aging. The first of the Baby Boomer generation (those born between 1946 and 1964) will turn 65 in 2011. It is estimated that 20% of the population will be 65 or older by 2030 [1].

Physiological changes associated with aging, such as decrements of sight, hearing, dexterity, motor functioning, hand-eye coordination and cognitive processing, make new screen technologies more difficult to use [2]. These changes occur regardless of gender, race or culture, and can have significant impact on quality of life. What used to be a simple trip to the grocery store may not be so simple for those experiencing these changes of aging.

The ability to read food and product information at the grocery store in a clear, concise and consistent format, is important to sustaining independence. Adults with dietary restrictions, such as the need for lower sodium intake, diabetes, and/or allergies, make it imperative to be able to read all nutritional information on packaging. This includes the nutritional facts, drug facts, supplement facts, as well as the complete list of ingredients. There is a lack of consistency in product packaging and not all information is visually presented in the same manner. This lack of consistency can create frustration among shoppers, in addition to the text sizes being too small for aging eyes to read, it can cause illness for someone with a dietary restriction that cannot properly read the product's contents.

The purpose of this project is to discover the proper design strategies for an application to be developed for a wireless device (i.e. Apple's iPad) to be used in a grocery retail environment. The device will function with Radio-Frequency Identification (RFID) tags placed on each product in the grocery store. The device will receive the product information and present it in a clear, concise and consistent

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 285–293, 2011.

format so shoppers will be able to read all of the nutrition facts and ingredients listed on a package.

During the second phase of this project, the following questions were answered with IRB approved usability testing:

1. Are adults age 46 and over receptive to using the Web-based application?
2. Will the application prove easy to navigate?

1.1 Demographic Analysis

According to Lee Rainie at Pew Internet & American Life Project, between the years 2000 and 2010, the number of older adults (aged 46 and over) using the Internet increased from 40% to 74%. Using the same comparisons, 63% have broadband at home, 81% own a cell phone, 46% connect to the Internet wirelessly and greater than 50% use cloud computing [3].

This group comprises those in the Baby Boomer generation (born 1946-1964), the Silent Generation (born 1930-1945), WWII children (born 1915-1930), and children of the Great Depression (born prior to 1915). While this group has enormity in numbers, it is not homogenous. Although similarities do exist, American society recognizes the life course theory through labeling blocks of years for any given population segment such as: young adult, adult, middle-aged adults, older adults, retired adults and very old adults [4]. While difficult, accepting these labels requires acknowledgment and realization of our advancing age in physical, psychological and social areas [5].

1.2 Design Concept

The older adult age group consists of varying socioeconomic levels and varying degrees of technological savvy. The concept was to develop an application for the grocery retail environment to address the issues of physiological changes that occur with aging when shopping at the grocery store. Accessing product information about food, supplements and drugs in a clear, concise and consistent format, was the main focus for developing the SCIP application.

Part of the concept is for the grocery store to provide a wireless device (i.e. Apple's iPad) mounted on a shopping cart with a closed, secure WiFi network. Differing socioeconomic groups of older adults would have access to the benefits of this device by the store providing it. A user could also use their own wireless device to login securely to the store's network. All shoppers of the grocery store of any age would be able to use the device, in addition to the older adult demographic that was researched for this project.

Each product in the store would be labeled with a RFID tag that the wireless device would read via a wireless antenna contained in the tag over the store's secure, closed WiFi network. RFID tags would complement barcodes and barcode readers, since they are more efficient, easier for devices to read and are capable of storing large amounts of information in their microchip. Nutritional information, pricing, expiration date, country of origin, size, weight, and further product information can be stored in the RFID tag related to a specific user's shopping needs.

Thomas Friedman, in his 2007 book *The World is Flat*, refers to RFID technology by stating, "This is clearly the wave of the future. RFID technology and sophisticated order analysis tools that monitor even the most minute market activity are rapidly leading us toward industry's holy grail—absolute balance in supply and demand," [6].

The mounting bracket on the cart would be adjustable to accommodate differing user heights and reduce glare, as well as allowing users with their own mobile devices to be able to mount them on the cart.

The team designed two different low-fidelity prototypes of user interfaces, User Interface 1 (UI1) and User Interface 2 (UI2), for the SCIP application. Both interfaces followed the same navigational flow and displayed the same types of information.

The following design guidelines were taken into account when the interfaces were designed [7]:

- Typeface: Use a sans serif typeface that is not condensed. Avoid the use of serif, novelty, and display typefaces.
- Type size: Use 12 point or 14 point type size for body text.
- Type weight: Use medium or bold face type.
- Capital and lowercase letters: Present body text in upper and lowercase letters. Use all capital letters and italics in headlines only. Reserve underlining for links.
- Physical spacing: Double space all body text.
- Justification: There are three ways to justify type; left, full, or centered justified. Left justified text is optimal for older adults.
- Color: Avoid yellow and blue and green in close proximity. These colors and juxtapositions are difficult for some older adults to discriminate. Ensure that text and graphics are understandable when viewed on a black and white monitor.
- Backgrounds: Use dark type or graphics against a light background, or white lettering on a black or dark-colored background. Avoid patterned backgrounds.

In regards to navigation through different screens, the team lessened the amount of vertical scrolling, kept the layout consistent, eliminated the need for a mouse by using a touch screen interface with a virtual keyboard, incorporated text with icons, used pull-down menus sparingly, and incorporated backward and forward navigation arrows on each page.

2 Evaluation Measures

Each of the five users were read the testing script prior to beginning the usability test. Each user test began with UI2, followed by UI1 with the same task list. After testing was completed, the team discovered that UI1 and UI2 should have been alternated as the beginning interface for each test. Instead, UI2 was used as the first interface for all tests. It is the author's hypothesis that the time spent per task on UI1 was less than it would have been if presented in an alternating fashion, due to the user's previous familiarity with the task list.

The user's mouse clicks and voice were recorded during each session and after both interface tests were completed, the user completed a questionnaire.

The first questionnaire was based on the user's demographic information. Questions included age, gender, native language, occupation, computer use, Web use, cell phone use, smart phone use, iPad use, online shopping, grocery shopping routine, problems reading product information on food product packaging and dietary restrictions. The questions were multiple choice.

The second and third questionnaires were identical, but one was related to UI1 and the other UI2. Questions were related to the interface's ease of use, accessiblity and usefulness. The questions were based on the Likert Scale and short answers.

2.1 User Testing

Survey participants were recruited using posters and word of mouth advertising methods. Software was used to record the screen to track mouse clicks and also record audio while the user thought aloud during testing. The recordings were tabulated to calculate where and how many mouse clicks were used during each user's testing session to determine errors.

The test administrator followed the testing procedure outlined below:

- Testing script was read verbatim and followed.
- Task list asking the participant to find nutrition facts for a product, find ingredients listed in a product, compare the amount of sodium in two products, add one product to shopping cart and checkout. This process was completed by the users using pre-filled form fields so that no private or financial information was collected.
- Survey questionnaire was completed by each user asking age, occupation, familiarity with technology, and post-test questions were given to determine participants' views on each of the two user interface designs application after completion of tasks.

2.2 Test Design

The team designed two different low-fidelity prototypes of user interfaces for the SCIP application. Both interfaces followed the same navigational flow and displayed the same types of information. However, the visual presentation of elements varied drastically. Prototype UI1 was based on using icons with text and prototype UI2 was a predominantly text-based navigation set within a modular grid format.

2.3 Testing Process

The two prototypes were tested during the same usability testing sessions with five users that were selected from a pool of respondents within the target demographic of adults age 46 and over. User demographics are displayed in Table 1.

Survey participants were recruited using posters and word of mouth advertising methods. The tests were conducted over a two-week period with the five users. The users were given a task list at the beginning of the session to use the Web-based application on a laptop computer where the screen was recorded to track mouse clicks, while simultaneously recording audio of the user thinking aloud.

Table 1. User demographics

User Demographic Facts Summary	
Question	%Breakdown
Age range	80% 46-55
Gender	60% female/40% male (5 users total)
Native language	60% English
Occupation	80% Professional
Computer use	100% Daily
Web use	100% Daily
Cell phone use	80% Daily
Use a smart phone	60%
Used iPad before	60%
Shop online	60%
Grocery shop routine	80% Weekly
Neither of the males in the study had dietary restrictions, but all three women had at least one. Two out of three women needed reduced sodium and reduced sugar intake. 60% of the users have problems reading the nutritional facts on food product packaging and 40% have problems reading the ingredients.	

It would have been ideal to have the application tested on a touch-screen mobile device, however, since the rapid prototyping method was used, the team created a low-fidelity application within a Web browser for use on a laptop computer.

3 Findings and Discussion

Both of the UI designs depend on the user already being registered in the SCIP program. The registration process can be completed at a store with the assistance of customer service or online via secure Web site. The user would fill out a submit form with their dietary restrictions so the SCIP application would be able to warn the user if any product in their shopping cart contains the ingredient(s) that is among their restrictions. For instance, if a user is allergic to peanuts, then the SCIP application would display a warning if that ingredient was present in the product when it was scanned.

The user would also be able to compare the nutritional facts with another product. For example, the sodium content per serving in two different kinds of canned soup could be compared to see which contains the lowest amount. This feature would be useful for someone keeping track of sodium intake.

In order to remove an item from the cart, the user simply places it back on the shelf as the device automatically removes it from the cart's contents. Once the user is ready to checkout, the checkout button is touched and the registered user's banking information or credit/debit card on file is charged for their purchases. This eliminates waiting in a checkout line to have the cashier scan each item and receive payment. The user can choose to bag their own groceries while shopping, or have a store employee bag their groceries upon checkout.

3.1 User Interface 1 (UI1)

UI1's design (Figure 1) was based on conventional Web site design. Contrast was kept between background and foreground to insure that older adults' reduced visual acuity was taken into account and the screen would remain readable. Type sizes were kept at 12 points minimum. Using the touch screen interface takes into account declining motor skills associated with aging, eliminating the need for a mouse or double-clicking. It also offers an alternative to the iPad's pinch-to-zoom option by keeping icons on the screen to adjust text sizes by touch.

Three options are presented at login.

1. Scan shopper card that is embedded with RFID tag
2. Use thumbprint
3. Use username and password

Fig. 1. Nutrition facts screen (left) and compare products screen (right) from UI1

According to User 3 the UI1 prototype was too time consuming and complex. User 3 mentioned confusion about the various login options. The user wasn't able to complete all tasks because the wrong menu path was selected which lead to a scramble of the task steps. The overall results of user testing were that 80% favored UI2 for both aesthetic design and functionality.

3.2 UI1 Errors

- Task 2.
  Scan a product.
  - 80% of the users did not immediately notice the scan product icon at the bottom of the screen. User 4 clicked 5 times on the wrong icons, as did user 3. User 2 and user 1 both clicked the wrong icon once before clicking on the correct one.

- Task 3.
  Find nutritional facts for a product.
  - 60% of the users had problems locating the nutritional facts. User 3 did not complete this task, even after assistance.
- Task 5.
  Compare the amount of sodium in a product with another product.
  - 60% of users had four or more errors while comparing products.
- Task 6.
  Add one of the products to your shopping cart.
  - User 1 said, “I am looking for add to cart button, by this time I would say screw this and get out of here. How do I get back?”
- Task 7.
  Scan another product.
  - 80% of the users had one or more errors while trying to scan another product.
  - User 1 said, “This one doesn’t work, this one doesn’t work, scan another item… far too small text.”
  - User 3 said, “I can’t read the text.” Even after adjusting her bifocals the user required assistance to find the scan-item and check-out button in the shopping cart.
- Task 8.
  Find ingredients for a product.
  - 80% of users had one or more errors while attempting to locate the ingredients.
- Task 9.
  Read the first ingredient in the ingredient list.
  - 60% of the users did not find the icon for the ingredients at the bottom of the screen. User 2 clicked 6 times on other icons before reaching the ingredients page.
- Task 10.
  View the total amount of sodium for the contents in your shopping cart.
  - 60% of users had one or more errors while looking for the total amount of sodium in the shopping cart.

3.3 User Interface 2 (UI2)

UI2’s design (Figure 2) was a novel interface concept based on the touch screen and inspired by the game Bingo. It utilizes large, square buttons in a modular grid structure.

Contrast was kept between background and foreground to insure that older adults’ reduced visual acuity was taken into account and the screen would remain readable. Type sizes were kept at 12 points minimum, with very minimal use of icons. Using the touch screen interface takes into account declining motor skills associated with aging, eliminating the need for a mouse or double-clicking.

Two options are presented at login.

1. Sign in as a registered user

2. Use as a guest, without logging in

This UI offers the option for the user to either login, based on previously registering with SCIP program or by simply using the system as a guest.

Fig. 2. Login screen (left) and nutrition facts screen (right) from UI2

3.4 UI2 Errors

- Task 3.
  Find nutritional facts for a product.
  - Users did not recognize the nutritional facts by the percentages that were shown, or the layout in squares. An overwhelming 80% of the users had at least one error on this task. User 2 repeatedly clicked on the words, “nutritional facts” in hopes that a new window or new page would appear with the information. The nutritional facts were already displayed on this page after task 2, which was to scan another product was completed.
- Task 5.
  Compare the amount of sodium in a product with another product.
  - 100% of the users completed with this task with errors. User 2 was looking for an option to create a list to make the comparison work. The user wanted the page to do something more. After clicking on product search for the second time, the user needed assistance on how to get to the correct page to perform the task.
  - User 4 liked the color coding for the compare product screen. Nevertheless, the user was confused that the same color’s (orange and grey) where used for other buttons like the compare product and nutritional facts button. Also the user was unclear about the chosen units for nutritional facts. User recommended to have no transparency for the “scan item” button on the side menu. The user noticed that the checkout buttons placement didn’t correspond with the overall layout of the interface.

4 Conclusions

Since 80% of the users preferred UI2, the team will focus on redesigning that interface. The errors and preferences in the design of UI2 were consistent among the users. Tasks three and five proved to be the two tasks that were difficult for almost all users.

An interesting correlation became apparent in terms of color preference and gender. The color preference for UI2, was 60% (all of the female users) found the color palette of orange and grey pleasing. The male users (40% or all of the male users) strongly disagreed with the color choices.

The color palette of blue and grey used in UI1, created a mixture of pleasing vs. not pleasing. Only one female user one male user found the colors attractive.
The team will also add a strong emotional connection to the user. This would encourage user participation with the application and create a more pleasurable shopping experience. The team will use a more bold color palette to create contrast and evoke an emotional response from the user. They will also consider giving the SCIP application an identity based on an avatar.

In addition to adding an emotional connection, the team will add haptic feedback to the interface. Using sounds and vibration will help the user feel that the interface is responding to their actions.

References

1. Czaja, Sara, J., Lee, Chin, C.: Designing computer systems for older adults. In: Jacko, Julie, A., Sears, A. (eds.) The Human-computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Lawrence Erlbaum Associates, New York (2001)
2. Selwyn, N., Gorard, S., Furlong, J., Madden, L.: Older adults' use of information and communications technology in everyday life. Ageing & Society 23, 561–582 (2003)
3. Rainie, L.: Baby boomers in the digital age. Pew Internet & American Life Project. In: Baby Boomer Summit – Ask the Analyst Conference Proceedings, March 19 (2010)
4. Steenblock, S.: User centered design evaluation of the grocery store environment: A study based on an aging population's perception of the person-environment fit. Thesis (M.A.), Iowa State University (2010)
5. Coburn, K.L., Lawrence-Treeger, M.: Letting go: A parent's guide to understanding the college years. Harper Collins Publishers, Inc., New York (1997)
6. Friedman, T.L.: The world is flat. Picador, New York (2007)
7. National Institute on Aging & National Library of Medicine, Making your Web site senior friendly (2004)

Part IV

HCI in Complex Environments

An Analytical Alarm Flood Reduction to Reduce Operator's Workload

Jens Folmer, Dorothea Pantförder, and Birgit Vogel-Heuser

Chair for Automation and Information Systems
Technische Universität München, Boltzmannstraße 15,
85748 Garching near Munich, Germany
{folmer,pantfoerder,vogel-heuser}@itm.tum.de

Abstract. In the domain of process control, an alarm flood is a situation when there are more alarms generated by the automation system than can be physically addressed by a single operator. To reduce alarm floods an analytical approach, so called AADA (Automatic Alarm Data Analyzer), has been developed to learn these alarm floods by itself. Finally, this behavior can be integrated into process-visualizations which illustrate only the root cause of an abnormal plant state. To increase the operator's awareness during abnormal plant states, a combined approach of the ADDA and the 3D process-visualization is presented in this paper. This approach has to reduce alarm floods and to display the most important information of a plant to the operator during runtime.

Keywords: 3D Visualization, alarm flood reduction.

1 Introduction

Process Control Systems (PCS) are one of the core elements in process control industries, e.g. gas oil refineries or pharmaceutical and chemical plants [1]. The PCS presents the most important information about the process plant to the operator. Beside the observation of process data, operators have to recognize abnormal plant states.

In general, the PCS's visualization is divided into two different parts closely related to each other. Operators get an overview by observing the visualization (see figure 1), which is based on P&ID-diagrams, line diagrams and tables to visualize process data and to show the plant's state based on graphical and textual elements. The graphical elements represent automation device of the plant, e.g. pump, valve, motor, pipelines, the tag information. The most important values to operate the process are visualized related to the elements, e.g. the actual value and the set value of the process. If the conditions for an abnormal plant state come true the graphical elements change the color, e.g. into yellow or red. Changing the color indicates an alert or an alarm (in the following the word notification is used for alarm and alerts), to call operator's awareness. Detailed information about the abnormal plant state is provided when operators have to switch from the process visualization to the alarm

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 297–306, 2011.

list. Usually, alarm lists are structured as a table where every standing and unaccepted notification is displayed in one row. Alarm systems ought to support the operator controlling potentially dangerous situations before the Emergency Shutdown System (ESD) is forced to intervene [1].

Thus, operators face two problems: The first one concern that the operator has to handle a vast quantity of notifications and it is difficult to manage those notifications [2]. There are notifications that occur due to causal dependencies between automation devices, e.g. pump, valve and sensor in one manufacturing line. Every automation device may raise a notification that is reported to the operator. Several standing notifications appear in the alarm list. This is one effect for alarm floods. During alarm floods operator just acknowledges the notifications while handling the abnormal situation regarding their process knowledge and experiences. In such a situation operators have not enough time to analyze every notification properly without exceeding the recommended operator response time. Furthermore, there is a risk that operators miss an important alarm [3]. Therefore, the reduction of alarm floods has a high priority when improving alarm systems design [1].

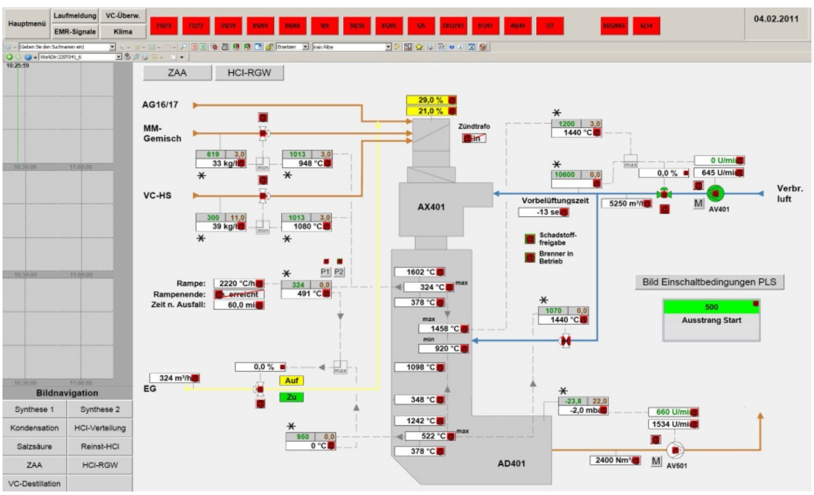


Fig. 1. 2D visualization of a process control plant

The second problem mention concerns the need of switching from the process-visualization to the alarm list to get detailed information about the raised notification. Since the process visualization is based on a graphical display and the alarm list is structured like a table. Hence, the operator loses sight of the process visualization by switching the visualization in order to appraise the alarm list. For instance, the operator is not going to recognize the difference of the actual and the responding set value becoming abnormal. Regarding the alarm list, the spatial information which automation device raised a notification or which one caused the abnormal plant state is lost. In that case, operators have to link information from the alarm list and the process visualization based on their cognitive performance.

Even today, there is no solution for alarm flood reduction during run-time of a plant integrated into the process visualization.

In this paper, we present a combined approach of 3D process-visualization and an integrated algorithm to reduce alarms floods. In section two the state of the art for analyzing data is presented and 3D visualization in section three. A new 3D visualization approach with an integrated alarm flood reduction is described in Section four. Finally, the discussion, the conclusion and the future works are presented in the section five respectively in the section six.

2 Learning Plant's Signal-Behavior

In the following the state of the art learning plant's signal-behavior during run-time is described.

For the alarm flood reduction Kurz [4] presents a manual approach to analyze alarm logs. Therefore, he focuses on the top alarms of a plant and tries to find out the main causes and effects between devices and alarms. Results are that five alarms are responsible for 87% of all alarms in one plant section. An analysis of the top 40 alarms in another plant shows that four alarms are the main causes which are responsible for disturbances. Actions to be taken are derived from the analysis to handle alarm floods and realizing the number of alarms being reduced from 1580 down to 440 alarms per day (72%).

Bauer [5] presents an approach for the probabilistic analysis of process data to detect causal relationships of process data generated in case of an error during the run-time of process plants. To classify process data, Bauer uses a decision tree to determine an appropriate calculation method for the analysis. Concerning the classification, the main important criterions for process data are:

- The quantity,
- The oscillation,
- Temporal linear dependency between process data,
- Temporal linear distribution of process data.

Based on the classification-criterions, calculation methods are suggested to select an appropriate calculation method, e.g. next neighborhood, transfer entropy or cross-correlation. The aim is to determine causes and effect's dependencies that are reasons for the faulty behavior of a plant. Bauer shows the causal dependencies by generating a cause-effect map.

To recognize error conditions of a plant, Klein [6] focuses on an automata-based approach to learn the behavior-model during plant's run-time by observing I/O-signals of I/O-Devices in every plant cycle. Therefore, Klein developed a new type of automata called NDAAO (non-deterministic autonomous automata with output) and a parametrizable identification-algorithm. Roth presents model-criterions which are used for the qualitative evaluation of the behavior's complexity (precision of the model) and the model's structure (size of the model). Both are important for an efficient calculation-time to recognizing plant's failure during plant's run-time. Referring to Klein's research, Roth [7] extended the automata approach. Roth has developed failure-indicators to detect differences between the learned set- and failure-behavior. Roth proposes integrating time-aspects in future works.

Process-mining is applied to find structures in unstructured event logs. Current research has focused on extracting a workflow-based process model. Kim [8] presents an FP-tree algorithm (frequent pattern tree) to detect frequent patterns stored in a log file. Before analyzing for frequent patterns it is needed to define a direct causal matrix to serve several purposes. Based on the direct causal matrix Kim is analyzing the log file by using a modified FP-tree for the process discovery. Usually Process-mining faces business-processes and is not established to analyze alarm logs in the domain of process control. Furthermore, this approach is not able to find causalities by itself.

2.1 Automatic Alarm Data Analyzer (AADA)

Regarding our research, we developed the automatic alarm data analyzer (AADA), an algorithm that is to find causal dependent notifications. The AADA analyzes notifications stored in alarm logs. An alarm log contains notifications recorded during the run-time of a plant. The aim of the AADA is to recognize causal dependent notification and generates alarm-sequences based on the alarm logs. Within alarm-sequences are causal dependent notifications to point out the root cause and their effects during an abnormal situation. Furthermore, we adapt the ADDA to except user inputs before the start of an analysis e.g. if the alarm log contains notifications generated by a fast or slow process. Additionally, the user can enter which plant sections have to be analyzed or what kind of notifications (alert, alarm etc.) have to be considered.

However, the alarm log contains notifications from all parts of a plant. Every generated notification is stored in the same alarm log. Depending on the size of a plant, notifications are stored temporal closely but are not dependent. Therefore, notifications generated by several independent plants could have a similar time stamps. By analyzing the alarm log these notifications appears next to each other and may be recognized as an alarm-sequence. This distorts the results of the analysis. Hence, generated sequences have to be verified by plant operators who know the process plant. Nevertheless, most of the alarm-sequences are relevant and are useful to improve the alarm management system to prevent design mistakes.

But a re-engineering of an alarm management system is time consuming. Hence, it is an advantage if the generated alarm-sequences, which are verified, could be integrated into the visualization for pre-processing the notification. Pre-processing by means of incoming notifications are analyzed by the AADA during run-time. Meanwhile, the AADA extracts the root cause of a recognized alarm-sequence by suppressing remaining messages concerning one alarm-sequence. Finally, only the root cause of an abnormal situation is indicated in the visualization.

3 3D Process Visualization

To display process data in machine and plant manufacturing 2D visualization is state of the art in process control. Current HMI systems use line diagrams, bar charts and tables to visualize process data or P&ID diagrams for structural information. These visualization elements are available in libraries and can be integrated into the process screen and be connected to process variables in a simple way. In some applications, e.g. tanks or piping are represented in 3D but from the operators' point of view, there is no additional benefit generated by this kind of 3D visualization. Even though 3D visualization is often considered as gimmick, there are different studies that show the benefits of 3D in process control. Beuthel [9] and Hoppe [10] proved the advantages of 3D process data visualizations for coal-fired power plants and electric power grits. They arranged 3D elements (e.g. bar charts) spatially on a 2D structure diagram (CAD, P&ID) to display actual process values. Both studies measured the reaction time and the processing time to handle problem situations in 2D as compared with 3D visualizations. The results showed an advantage for 3D representation.

The general applicableness of 3D process data visualization and their interaction possibilities for process data visualization and data analysis of multidimensional data fields was evaluated by the authors for typical applications of industrial process control [11]. For example a 3D visualization for monitoring temperatures in an ethylene reactor was provided [12] and implemented in the plant's process visualization.

The time consuming development could be a reason for the rarely use of 3D in process control. Each 3D scene has to be programmed step by step. For this reason, we developed a prototype library with 3D objects, so called 3D pattern [13]. These 3D patterns use standard interfaces like OPC or ActiveX and can be integrated in most state-of-the art HMI systems as easy as 2D visualization objects. The library was developed based on real application examples from industry and has been evaluated by experts.

The benefit of 3D in process data visualization was evaluated empirically by the authors by means of the application example of a continuous hydraulic press. 2D visualization, 3D visualization and 3D visualization with interaction of process data as well as different training situations were evaluated comparatively. The results of the experiment showed a significant advantage of 3D with interaction in error detection of complex problem situations [14] [11]. This result corresponds with Wickens' "proximity-Compatibility"- principle (PCP) [15]. This principle states that tasks that require the integration of information, benefit from a perceptual proximity of the display or visualization. The integration of information is necessary in particular for the detection of complex problems. These problems require the combination of three or more process values for the analysis of the system status. At the same time also a minor subjective factor of stress could be proven during the process control in the group of 3D with interaction. The fact that these only could showed in the group 3D with interaction, suggests that interaction with the 3D scene has a substantial influence on the information reception.

4 3D Visualization Combined with Alarm Flood Reduction

4.1 Visualization for Cause-Effect Disturbances

There are two main objectives for the new visualization approach. The first one concerns to the integration of cause-effect dependencies into the visualization. The second one faces the problem to visualize only the root-cause by proper elements.

In contrast to 2D, the 3D visualization offers another dimension to present more information to the operator in contrast to 2D. Concerning Wickens' PCP we use the third dimension to couple the information of process data, notifications and structural information in one process screen.

For the 3D visualization cylinders are used as the basic element shown in figure 2. This element is divided in several parts. The symbol for process control objects (PCO) is placed on top of the graphical element. The PCO-symbol is similar to the standardization ISO 14617 [7] as it is used for 2D process-visualizations. PCO-Symbol is placed on top of the cylinder (1) given in figure 2. The circle around PCO-symbol shows the percentage MTBF-value (2) which represents the maximum operating time of the corresponding automation device.

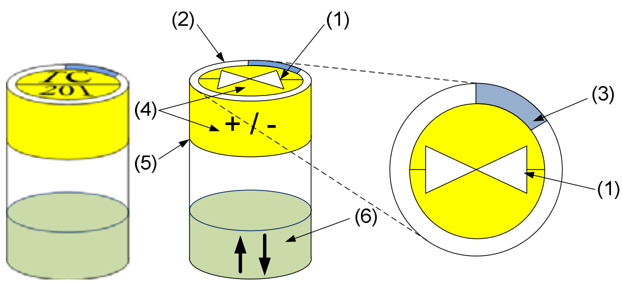


Fig. 2. 3D-PCO (valve) with process-, MTBF-value and alarm/warning

Concerning the colored part of the MTBF-circle (3), it represents the actual operating time of the device. In case those devices exceed the maximum operating time the MTBF-circle is fully colored (3). The MTBF-value is determined as the predicted elapsed time between inherent failures of a system during operation. Moreover, the MTBF-value is important for safety critical systems. The MTBF-circle tells the operator whether a malfunction of an automation device could be the cause of an abnormal plant or not, e.g. incorrect set up of values. The MTBF-circle is also useful to notifying the operator by disclosing necessary forthcoming maintenance, to prevent plant breakdowns. Often, the automation devices' MTBF-value is defined in the data sheet of the manufacturer and can be integrated into the visualization by calculating and counting the expected time for a failure of the responsible automation devices. However, the combination of the MTBF-value and the related alert/alarm helps the operator to better assess the abnormal situation.

When certain conditions creating notifications comes true, the upper part of the 3D-PCO starts blinking changing between white and the corresponding color for alert or alarm (4). To indicate an alarm high or an alarm low a plus "+" or respectively "-" placed on the upper part of the 3D-PCO. In case that the alert/alarm causing condition has gone and the operator did not acknowledge the notification, the blinking stops but the color is still displayed. The color disappears as soon as the operator acknowledges the notification.

Hence, if an abnormal situation occurs and an alarm/alert is illustrated the operator might click on the corresponding 3D-PCO to get detailed information about that notification. A faceplate like in common process control systems is illustrated and the common faceplate including the alarm sequence list is displayed. Thus, the faceplate supports operators assess abnormal situations.

Another important type of values to operate a plant are process values. The difference between the actual and the set value is displayed in the lower part of the 3D-PCOs (6). The arrow indicates whether the process value is ascending or descending. For an ascending process value, the arrowhead points from bottom to top while a descending one is displayed by the arrow pointing from top to bottom.

Concerning the 3D-PCO, the difference between the actual and set value is displayed. The maximum displayed difference (5) is determined as the maximal strength of the device concerning the value-differences. The maximal strength of all devices is defined as the technical system by Lauber [17]. The technical system is determined as the whole technical equipment realizing the process control. In case that the difference reaches the maximum value (5) the automation device could be damaged. If there is no difference between the process values the height of the process bar is zero, thus, part (6) is empty.

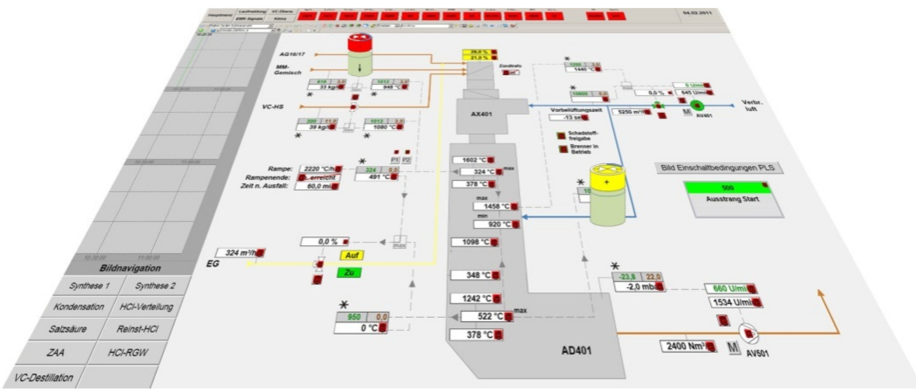


Fig. 3. 3D-PCO (valve) with process-, MTBF-value and alarm/warning

To change the focus within the visualization the operator can interact with it. By zooming in and out, the operator focuses on a plant section in the visualization, e.g. to have a proper view on the MTBF-value or the change process values of the PCOs. Due to the height of the 3D-PCOs, other graphical or textual elements could be placed behind other ones. Hence, operators have to rotate the visualization to avoid covering elements they wants to focus on. By using the bird's eye perspective the PCO placed on the visualization looks like the commonly applied 2D visualization but is additionally including the MTBF-circle around the PCOs.

4.2 Integrating Alarm-Sequences Generated by the AADA

The alarm sequences, generated by the AADA, contain causal dependent notifications. The first notification in each sequence corresponds to the root cause of an abnormal situation. Remaining notifications in an alarm sequence are effects of the root cause. To call operator's attention on the main problem in abnormal situations, pointing out the root cause of disturbances is necessary. Furthermore, to increase transparency and legibility of the 3D visualization, it is recommended to avoid placing a 3D-PCO on every existing PCO in the visualization. Integrating the alarm sequences into the visualization resolves these problems during run-time of process plants by suppressing the remaining notifications an just showing the most important notifications on the visualization. Furthermore, based on the alarm sequences, alarm floods can be reduced in the same way.

The moment, when an abnormal situation occurs, the AADA algorithm starts analyzing the incoming notifications. If an alarm sequence is recognized the AADA sends an event to the 3D visualization to trigger the 3D-PCO of the root cause. In case that there are notifications not recognized by the AADA as a part of an alarm sequence, the AADA designate the incoming notification as unrecognized. The 3D visualization will indicate every notification by a single 3D-PCO.

5 Discussion

The 3D visualization has been evaluated by experts of one of the global players in the chemical industry.

Concerning the evaluation, it has been pointed out some promising results. Visualizing the operation time (MTBF-value) of automations device could be very interesting for operators. Generally, the operation time is only available for asset management, especially for safety critical devices. The number of safety critical systems is not negligible. Depending on the process plants, the number of safety critical devices is between 15% and 70% in relation to the overall applied automation devices. Nowadays, operation times are not integrated into plant's visualizations but they might be an advantage. Operators are given information to assume whether an abnormal situation is based on malfunctions of automation devices or on other reasons. Furthermore, operators are able to anticipate abnormal behavior by considering the operating times of the 3D-PCOs. In fact, the pollution of automation devices might influence their operation time, e.g. when they come into contact with chemical materials, such as acid, base, fire, etc. In that case, the illustrated operation times may differ from the expected one.

The expert's evaluation points out that the actual deviation of process values relating to their set values is often more important than the actual process value of an automation device itself.. Furthermore, the illustration points out when the maximum of the allowed deviation is reached. Hence, illustrating the deviation of process values might call operator's attention rather than display actual values.

Alarm floods are problems in several process plants. The AADA itself have been used to analyze alarm logs recorded by five different process plants. Most of the frequent alarm sequences are significant but some alarm sequences are coincidental in the results. Hence, the alarm sequences have to be reviewed to apply the alarm sequences into the run-time visualization. Due to the fact that the AADA analyzes incoming notifications and interferes with the process visualizations, it is important that there is no erroneous alarm sequences integrated into the visualization

Another point is concerning the system integration of the alarm-sequences into the visualization. Regarding the expert's statement, it is important that the calculation time for analyzing incoming notifications do not have to influence the overall process, the speed of the visualization and the data transmission speed of networks.

6 Conclusion and Future Works

Concerning call operators' attention and for alarm flood reduction during the run-time of plant, a new 3D visualization with interaction is presented in this paper. The visualization is based on P&ID diagrams. The 3D visualization includes cylindrical process control objects (3D-POC), which are divided into a MTBF-value part, an alert/alarm part and a process value part. Hence, operators have the most important information in sight. Regarding the interaction, the operators are able to zoom in and to rotate the visualization. Therefore, the operators can focus on selected process sections or uncover process sections that are placed behind 3D-elements.

For alarm flood reduction alarm sequences are integrated into the visualization. The alarm sequences are found by an automatic alarm data analyzer (AADA) algorithm which analyzes alarms and alerts stored in alarm logs. These alarm sequences contain causal dependent alarms and alerts whereby the first alarm or alert of a sequence represents the root cause. By integrating these alarm sequences into the 3D visualization and suppressing the remaining notifications during the run-time of a plant, only the root cause is display to the operator.

Future works will focus on how to enhance the analysis based on AADA algorithm to avoid improving the alarm sequences by a plant operator. The benefit will be an AADA finding more significant alarm sequences without investigating the sequences by a plant operator.

Furthermore, we will implement the combined approach of 3D visualization and alarm sequences into the visualization of a plant-demonstrator. The aim is to measure the automation network to compare the transmission time by using the new visualization approach related to the common process control visualization.

Thereafter, we will implement the 3D approach into the visualization of a plant simulation. The plant simulation is based on real processes of plants and is used for the education of operators in the chemical industry.

Additionally, we will design an experimental environment for the empirical evaluation. Primary, a preliminary experiment with students will carried out. Thereafter, we will realize the main empirical evaluation. We expect that the planned empirical evaluation will point out that the 3D alarm visualization approach achieves better results compared to 2D. Operators will be able to recognize a complex problem quicker in contrast to the traditional 2D visualization and the alarm list approach.

References

1. EEMUA Publication 191: Alarm Systems – A Guide to Design, Management and Procurement, 2nd edn. (2007)
2. Bransby, M. L., Jenkinson, J.: Alarm management in the chemical and power industries - a survey for the HSE. IEE Colloquium on Stemming the Alarm Flood, pp. 1–8 (1997)

3. Brown, D. C., O'Donnell, M.: Too much of a good thing? - alarm management experience in BP Oil. 1. Generic problems with DCS alarm systems. IEE Colloquium on Stemming the Alarm Flood, pp. 5/1 –5/6 (1997)
4. Kurz, H.: Alarmmanagement – Ziele, Erfahrungen, Nutzen. Automatisierungstechnische Praxis (atp) 2, 86–90 (2008)
5. Bauer, M.: Data-Driven Methods for Process Analysis. Dissertation University College, London (2005)
6. Klein, S.: Identification of Discrete Event Systems for Fault Detection Purposes. Dissertation Technical University of Kaiserslautern (2005)
7. Roth, M.: Identification and Fault Diagnosis of Industrial Closed-Loop Discrete Event System. Dissertation Technical University of Kaiserslautern (2010)
8. Kim, G.-W., Lee, S.H., Kim, J.H., Son, J.H.: An Effective Algorithm for Business Process Mining Based on Modified FP-Tree Algorithm. In: Second International Conference on Communication Software and Networks, ICCSN 2010, pp. 119–123 (2010)
9. Beuthel, C.: Dreidimensionale Prozessvisualisierung zum Führen technischer Anlagen am Beispiel eines Kohlekraftwerkes. PhD Thesis University of Clausthal-Zellerfeld. Papierflieger Verlag, Clausthal-Zellerfeld (1997)
10. Hoppe, S.M., Essenberg, G.R., Wiegmann, D.A., Overbye, T.J.: Three-dimensional displays as an effective visualization technique for power systems monitoring and control. Technical Report. University of Illinois, US (2004)
11. Pantförder, D., Vogel-Heuser, B.: Benefit and evaluation of interactive 3D process data visualization in operator training of plant manufacturing industry. In: 2009 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Antonio, USA, October 11-14 (2009)
12. Vogel-Heuser, B., Zeipelt, R.: Nutzen der 3D-Prozessdatenvisualisierung in der industriellen Prozessführung. atp - Automatisierungstechnische Praxis 45(3), 45–50 (2003)
13. Friedrich, D., Moritz, S., Pantförder, D., Vogel-Heuser, B.: 3D-Patterns for the Data Analysis and Operation of Process Plant. In: International Association of Science and Technology for Development- Human Computer Interaction (IASTED-HCI), Phoenix (2005)
14. Pantförder, D., Vogel-Heuser, B., Schweizer, K.: Benefit and evaluation of interactive 3D process data visualization for the presentation of complex problems. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 869–878. Springer, Heidelberg (2009)
15. Wickens, C.D., Andre, A.D.: Proximity Compatibility and Information Display: Effects of Color, Space, and Objectness on Information Integration. Human Factors 32(1), 61–77 (1990)
16. ANSI/ISA S5.1: Instrumentation Symbols and Identification. Standardization (2009)
17. Lauber, R., Göhner, P.: Prozessautomatisierung, vol. 1. Springer, Berlin (1999)

Self Replicating Robotic Strategies as a Catalyst for Autonomous Architectural Construction

Michael A. Fox

California State Polytechnic University, Pomona, CA, 91768
michaelfox@csupomona.edu

Abstract. The research explores examining future trends in robotics and how they can be applied to spatial interactive architectural environments. The strategy of using modular robotics of architectural space-making demonstrates an architecture whereby adaptation becomes much more holistic and operates at a very small scale. The strategy of using self replicating strategies as a catalyst for autonomous architectural construction was very much driven by the premise of an advanced architectural design studio. This paper highlights conceptual contributions by architecture students for alternative means of Martian Colonization through means creating architecture that creates itself. The parameters of the design project had three primary considerations including: The actual trajectory issues (how to get materials to the Mars), Chemical Processing (how to make materials on the Mars) and Space Manufacturing (how to fabricate and assemble/construct things on Mars). Of these central issues explored in this studio, the focus was primarily on Manufacturing as a process carried out by small modular robotics. The premise of the approach is that rather than sending a constructed architecture to space, we send tiny robotic modules that are capable of mobility and reproduction through automated fabrication techniques using in-situ materials. The modules with embedded sensors, self-healing composites, and responsive materials were designed to construct buildings aimed at adaptation. Such buildings could potentially respond in a humanlike way to counteract loads, reduce material and allow for active environmental adaptation. When enough of architecture of the colony has constructed itself – we send humans to inhabit it.Several examples by architecture students are highlighted whereby individual modules were created within the context of a space architecture design studio and applied to scenarios of space making at various scales. The design context primarily focused on the master plan of a colony for 10,000 inhabitants. The colony is an assembly of numerous discrete yet interconnected projects that include residential, public, civil, industrial, commercial, research, healthcare, and farming etc. The environment on mars was also seriously considered including: gravity, pressure, radiation, and the mass balance of resources and waste required for sustaining human life at such a scale. Students worked in teams of two to produce complete colony designs including the detailed development and a construction/ fabrication concept for one of the buildings. Students developed scaled prototypes of the system that successfully demonstrated the robotic aspects of their project. Physical models demonstrated actual robotics, structure and materials. Biomimetic strategies were employed as a means to satisfy adaptability in terms of form, processes, and systems. Central to biomimicry

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 307–317, 2011.

within the context of the work was an understanding of the process by which organisms grow and develop including includes growth, differentiation, and morphogenesis. In terms of adaptation, the area of morphogenesis was primarily studied as a means to create an architecture that ensures a continuous turnover of cellular-like robotic modules that dynamically ensure mechanical integrity similar to that of a living, evolving system.

The projects successfully demonstrate various strategies for mechanical design, locomotion and control.

Keywords: Modular Robotics, Architectural Construction, Space Architecture, Habitat Construction, Robotic Construction, Interactive Architecture.

1 Introduction

The strategy of using self replicating robotic strategies as a catalyst for autonomous architectural construction was very much driven by the premise of the advanced topic architectural design studio. Space Architecture is an ideal test bed to explore many of the basic interfaces which concern humans in a different way and that puts the subject of architectural construction into a different perspective. This particular was taught with collaborators from NASA JPL, JSC and Boeing and focused on automation and robotics in construction.

The focus of the design studio was to allow architecture students to make contributions to the conceptual design for alternative means of Habitat Construction on Mars through a means creating architecture that creates itself. The design project had three Primary Considerations including: The actual trajectory issues (how to get materials to the mars), Chemical Processing (how to make materials on the mars) and Space Manufacturing (how to fabricate and assemble/construct things on the mars). Of the central issues explored in this studio, research concentrated on Manufacturing as a process carried out by small modular robotics. Issues of Space Manufacturing, which is very much tied to chemical processing, therefore became paramount as carried out by small modular robotics at a variety of scales. Such systems were seen to demonstrate value for three primary reasons: 1) the required quantity of structural material resources is far in excess of what can be sensibly be launched from the Earth, 2) the required civil and structural engineering tasks dictate machinery requirements far in excess of what can be sensibly launched from Earth, and 3) the requirement of fabricating the components and building and maintaining the facilities.

The premise of the approach was that rather than sending a constructed architecture to space, we send tiny robotic modules that are capable of reproducing through automated fabrication techniques using in-situ materials. The modules with embedded sensors, self-healing composites, and responsive materials will construct buildings aimed at adaptation. Such constructed architecture could then also have the potential to respond to counteract loads and reduce material, change shape to block sunlight, allow for active ventilation and insulation, and prevent their own degradation. The general premise is that when enough of architecture of the habitat has constructed itself, we safely send humans to inhabit it.

2 Project Overview

The overall architectural designs primarily focused on the master plan of a self-sustaining settlement for 10,000 inhabitants. The settlement is an assembly of numerous discrete yet interconnected projects that include residential, public, civil, industrial, commercial, research, healthcare, and farming etc.

Students were required to develop hyper-efficient urban planning and architectural design which includes human and environmental interactions, sociology and psychology. Students worked either individually or in teams of two to produce complete designs of one component facility in the settlement including the detailed development and a construction/ fabrication concept. Students developed scaled prototypes of their systems that were required to demonstrate the robotic aspects of the project as integrated into the designs to optimize the performative aspects in terms of energy, mobility and robustness. Physical models were developed to simulate actual robotics, structure and materials. Issues of embedded computational control structures, communication and kinetic engineering were therefore paramount. The environment on the mars was seriously considered including: gravity, low atmospheric pressure, dust, radiation, and the mass balance of resources and waste required for sustaining human life at such a scale.

3 The Architectural Value of Modular Robotics

The idea of architectural building blocks with autonomous reconfigurable robotics is at the forefront of architectural robotics today. Designers in robotics are moving away from traditional (pick and place) uses of automated mechanical devices in architecture to transformable systems that are made up of a number of small robots. For many terrestrial applications ranging from cleaning carpets and windows to adjustable furniture, we are seeing a distancing from the precedent of figural humanoid robots to transformable discrete systems. The manufacturing technologies compounded with recent advancements in software (computational intelligence) for these systems allow robots to be increasingly smaller and smarter. Current advancements in evolutionary and self-assembling robots, specifically dealing with the scale of the building block and the amount of intelligent responsiveness that can be embedded in such modules, are setting new standards for robotics.

As architects and designers familiarize themselves with more diverse, responsive, and autonomous robotic systems, they are beginning to understand ways to apply them to dynamic situational activities and build them into systems that make up architectural space. Our furniture and entire spaces might someday be comprised of a multitude of interconnected assemblies of robotic modules that can reconfigure themselves for a variety of needs or desires. There are many important lessons to be learned in both distributed computation and small-scale robotics that can feed into a future paradigm of architectural space-making.

3.1 Performance Parameters and Precedent in Modular Robotics

Students in the context of this studio therefore needed to simultaneously consider the methods of movement, connection, geometry, and embedded intelligence of such small-scale robotic modules. In addressing the performance parameters of modular robotic design, concepts focused on several key strategies: 1) geometry 2) movement 3) connection 4) scale 5) materiality, and 5) embedded intelligence. The final objective of the approach was to create innovative designs that are minimally functional with the capability for evolving additional multi-functionality. An additional primary consideration is how modules can connect to each other with sufficient mechanical strength and then disconnect easily again without using too much energy. In addition to the tectonic objectives of the robotics listed above there are several architectural objectives that this project explores 1) It served as a vehicle for developing strategies for decentralized control dictating how individual parts of a collective system should behave and how local interactions between individual modules work in terms of forming structures and figuring out how to move them around. 2) It served to demonstrate the possibilities of architectural space-making with unprecedented levels of customization and adaptability.

The studio began with an inclusive overview of modular autonomous robotics that had architectural applicability including: The Biologically Inspired Robotics Group (BIRG) at the Swiss Federal Institute of Technology.[1] Modular reconfigurable robotics at the scale of furniture was also presented as developed at the Self-organizing System Research Group at Harvard University. Also, researchers at Caltech are developing robots made up of modular parts that work as a system to interpret and act upon information.[2] Hod Lipson and other scientists at the Cornell Computational Synthesis Lab have also begun developing multiple types of modular reconfigurable robots and evolutionary robots. These self-replicating prototypes were designed to allow for each object to be able to attach, detach, and reattach to different self-similar faces based on predetermined computational logic. These modular objects are able to connect to each other through electromagnetic connections, and the entire system has the ability to change its physical shape based on how it is programmed.

3.2 Component Module Design

The students began with a module design that geometrically satisfied the required performance parameters listed above in section 3.1. In these robotic modules developed by the students, the scale of the module was typically based on the size of the microprocessor board, battery, and mechanical parts that had to fit within each module. The architectural students' robotic explorations were limited by the current possibilities of manufacturing and of the inherent physical mechanics. The robotics were developed within the context of the studio and included an overview of electronics and tutorials of Audrino. Students developed a cursory understanding of basic circuitry and were able to apply a number of different sensors to their models.

For the most part the robotics was limited to a single module and that informed the parameters of the larger physical model. In most cases, designs were limited to practical extrapolations based on the development of a single module. Also informing

[1] Roombots. http://birg.epfl.ch/page65721.html

[2] Chih-Han Yu, Francois-Xavier Willems, Donald Ingber and Radhika Nagpal. "Self-Organization of Environmentally-Adaptive Shapes on a Modular Robot." Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, 2007.

the module development was a required diagram of their larger architectural strategy for in-situ resource allocation, fabrication and construction as well as how it can deal with water, air, and electrical. In the examples from the studio below, the students designed (and modelled) conceptually self-replicating models which would allow for each object to be able to attach, detach, and reconfigure according to predetermined computational logic. The projects successfully demonstrated various strategies for mechanical design, locomotion and control.

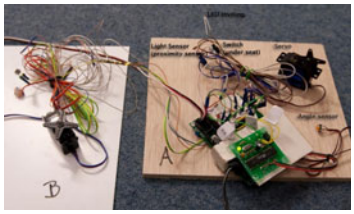


Fig. 1. Typical Robotic Module Electronic Components. Students used the Arduino developer kit and various sensors and actuators

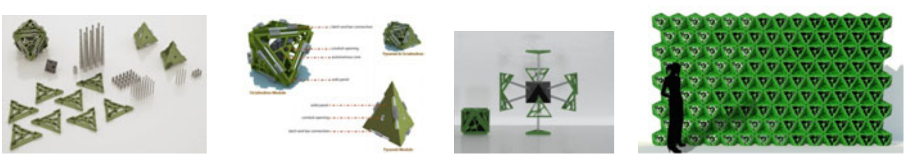


Fig. 2. This module used a panelized design with linear actuators for motion built around a dodecahedron hub

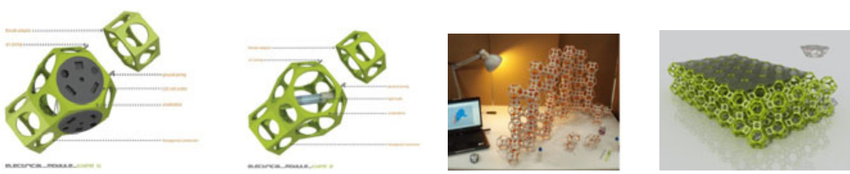


Fig. 3. This module used a male-female latching strategy and two separate complimentary modules

Fig. 4. This module used a six-pointed star with linear actuators on extending arms for motion and latching and a strategy for injection infill

Fig. 5. This module used a pendulum-based spherical strategy for motion which adhered with Velcro

3.3 Discrete Mechanical Assemblies and Decentralized Control

Understanding how to create architecture with autonomous reconfigurable robotics necessitates a clear understanding of a control structure. The important point is that each individual actuating module is controlled by a decentralized controller at a local level. This model of decentralized identification and control is based on neural networks and simplifies the implementation of the control algorithm. Decentralization is valuable on a number of points. In creating many self-similar parts, there is a redundancy in terms of control, an economic savings in terms of mass-production and an increased robustness to failure, in that if any single part fails, the system as a whole does not fail. When there are many unknown stimuli such as an exterior environment which is constantly changing, then decentralized intelligence is the obvious choice for being an effective way to handle the sensing and response (perception and action). Architectural robotics in a very general sense is built on the convergence of embedded computation (intelligence) and a physical counterpart (kinetics) that satisfies adaptation within the contextual framework of environmental interaction. The individual modules therefore can have a remarkable ability to communicate with each other even while being specifically task oriented. Decentralization then is a powerful control strategy for such systems of individually networked devices (in this case) whereby there is no central control system.

Most architectural applications are neither self-organizing nor do they have higher-level intelligence functions of heuristic and symbolic decision-making abilities. Most applications do, however, exhibit a behavior based on low-level intelligence functions of automatic response and communication. When a large architectural element is responding to a single factor then a centralized system can be effective in executing a command to a single agent, but when there are many unknown stimuli, or many small autonomous parts, then decentralized intelligence is the most effective way to handle the sensing and response. The more decentralized a system is, the more it relies on lateral relationships, and the less it can rely on overall commands. In a decentralized system there is normally no centralized control structure dictating how individual parts of a system should behave, local interactions between discrete systems therefore often lead to the emergence of global behavior. An emergent behavior can occur when a number of simple systems operate in an environment that forms more complex behaviors as a collective. The rules of response can be very simple and the rules for interaction between each system can be very simple but the combination can produce interactions that become emergent and very difficult to predict.

4 Biomimetics as Process Inspiration

As the overall designs necessitated a certain degree of dynamic response, biomimetic case studies were explored as a means to satisfy adaptability in terms of form, processes, and systems. Biomimetics studies systems, processes, and models in nature, and then imitates them to solve human problems. It lies at the intersection of design, biology, and computation. Put simply, nature is the largest laboratory that ever existed and ever will. Central to biomimicry within the context of this studio was an understanding of the process by which organisms grow and develop. This area of developmental biology includes growth, differentiation, and morphogenesis. In terms of adaptation, the area of morphogenesis was primarily studied as a means to create an architecture that ensures a continuous turnover of cellular-like robotic modules that dynamically ensure mechanical integrity similar to that of a living, evolving system.

Students performed case studies in modular autonomous robotics that had the potential to reproduce themselves. New available technologies like the fab@home 3-d printer which has the capacity to print with a wide palette of materials and mobile CNC routing robots became the inspiration for what might be possible architecturally with modular robotics. With the possibilities of such new CNC processes, students began to look at precedent in nature that was in-line with the adaptive issues that they were trying to solve. The heuristic approach is very bottom-up, in that you first design the brick (robotic module) and then the architectural possibilities are very much influenced by the inherent possibilities and limitations of that particular module. These modules then reference precedent in nature as an inspiration for how they could adapt.

The approach concerning biomimetics was to design architectural systems that could operate like an organism, directly analogous with the underlying design process of nature. Architectural robotics utilized at such a level could allow buildings to become adaptive much more holistically and naturally on a number of levels. Understanding the processes by which organisms grow, develop and reproduce then became an invaluable precedent for how such small mechanisms in an architectural environment could potentially operate.

5 Self Replication Strategies and ISRU

In terms of self-replication, students did a large amount of research on Martian in-situ materials resource utilization (ISRU) that could mined and used as building materials. There is a great wealth of materials that span the gamut of a necessary architectural palate. Numerous types of masonry bricks can be constructed using regolith[3]. As such a material has a low tensile strength and because views would be desired, the best strategy was to use masonry in combination with other materials such as glass

[3] Mackenzie, Bruce, "Building Mars Habitats Using Local Materials" pg 575 in The Case for Mars III: Strategies for Exploration. Stoker, Carol ed., American Astronautical Society: Science & Technology Series v74, 1987.

which could also be fabricated with regolith materia[4]. In light of the material possibilities, the studio then looked as existing precedent such as the Fab@Home 3-d printer which allows for the creation of multi-material objects. This allows a printed object to be functional rather than static. The Fab@Home has the ability to print conductive silicone, epoxy and cement. By mixing these materials into interesting geometries, fluidic and electrical conduits have been printed. This enables the creation of structural bricks that contain an edifice's plumbing and wiring pre-installed. These systems can be seamlessly intertwined with conventional building systems[5]. Students then developed different scenarios which incorporated bricks which could also create other bricks. Each project developed ideas of fabrication and mobility so that the some of the bricks became miners, some fabricators and some static bricks.

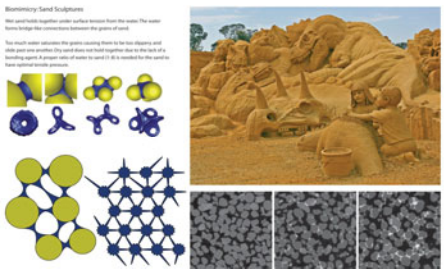


Fig. 6. Biomimetic Module, This module was articulated using the adhering properties of sand as a precedent for clustering

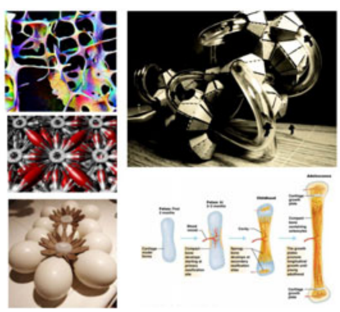


Fig. 7. Biomimetic Module. This module was articulated using the remodeling properties of bone as a precedent for growth.

[4] Lansdorp, Bas; von Bengtson, Kristian (2009). Mars Habitat Using Locally Produced Materials. In A. S. Howe, B. Sherwood (Eds.), Out of This World: The New Field of Space Architecture (Chapter 23, p. 311-315). Reston, Virginia, USA: American Institute of Aeronautics and Astronautics.

[5] H. Lipson, "Principles of modularity, regularity, and hierarchy for scalable systems." Journal of Biological Physics and Chemistry 7, no. 4 (2007): 125–128.

An important aspect here is that such building systems reposition the role of the designer. As Gordon Pask states in his foreword to the book, An Evolutionary Architecture: “The role of the architect here, I think, is not so much to design a building or city as to catalyze them: to act that they may evolve.” [6]

6 Creating Architectural Space

A primary goal of this studio was to make conceptual contributions to architectural systems that are made up of a number of small robots. In other words, students were required to conceive of futuristic architectural possibilities of this new direction in robotics. Manufacturing technologies compounded with recent advancements in software (computational intelligence) allow the robotic parts in these systems to be increasingly smaller and smarter. Current manufacturing technologies have allowed microprocessors to grow increasingly smaller, cheaper, and more powerful and we are seeing that we now have the potential to think of space itself as being organized in a computational network. These new standards are extremely exciting in light of the role of autocatalytic processes, defined here as a reaction product itself being the catalyst for its own reaction. In the context of modular reconfigurable robotics such processes describe how the pace of technological change is accelerating because of these processes. In other words, the process is “autocatalytic” in that smart, articulate machines are helping to build even smarter, more articulate ones. In the examples

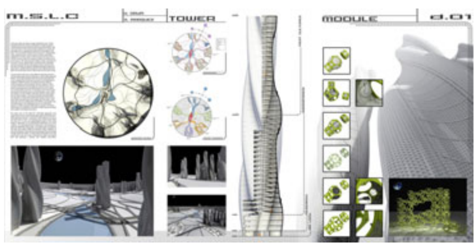


Fig. 8. Architectural Strategy Based on Module Design. These structures were developed with the module in Fig.3.

Fig. 9. Architectural Strategy Based on Module Design. These structure were developed with the module in Fig.4.

---

[6] Introduction to: An Evolutionary Architecture, by J. Frazer (London: Architectural Association Publications, Themes VII, John Frazer and the Architectural Association, 1995).

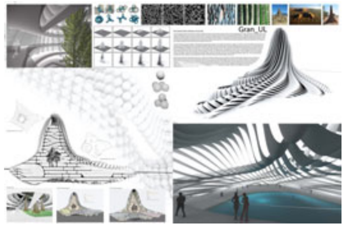


Fig. 10. Architectural Strategy Based on Module Design. These structure were developed with the biomimetic strategy in Fig 6.

Fig. 11. Architectural Strategy Based on Module Design. These structure were developed with the biomimetic strategy in Fig 6.

below, the students used their self-replicating module designs to construct the larger architectural environments. The projects successfully demonstrated various strategies for spatial adjacencies and relative scale. The designs also included site-planning of all buildings, transportation, access, emergency egress, and landscaping.

7 Conclusions

The research explores examining future trends in robotics and how they can be applied to spatial interactive architectural environments. The strategy of using modular robotics of architectural space-making demonstrates an architecture whereby adaptation becomes much more holistic and operates at a very small scale. This paper highlights conceptual contributions by architecture students for alternative means of Martian Colonization through means creating architecture that creates itself. This design project examines the value of self-reconfigurable robotics as basic architectural building blocks. The work was carried out in the context of an advanced topic architectural design studio. The design studio was successful in allowing architecture students to make contributions to the conceptual design for alternative means of settlement scenarios through means creating architecture that creates itself. The design project was carried out with three Primary Considerations including: The actual trajectory issues (how to get materials to the moon), Chemical Processing (how to make materials on the moon) and Space Manufacturing (how to fabricate and assemble/construct things on the moon). Of the central issues explored in this studio, the primary consideration was on Manufacturing. Several examples were highlighted whereby individual modules were created and applied to scenarios of space making at various scales. For architectural students, this was a highly research–intensive studio both in terms of Space issues and also robotics. Although the robotic aspects were very underdeveloped, the projects successfully demonstrated various strategies for mechanical design, locomotion and control. Biomimetic strategies were employed as a means to satisfy adaptability in terms of processes and systems that focused on growth scenarios that integrated self-replication.

We believe that this cursory exploration into architectural building blocks with modular autonomous robotics has great potential in space architectural applications. We also believe that there is a great amount of work to be done particularly in terms of scaling issues and in biomimetics. Such an extrapolation of advancements in both robotics and new materials demonstrates an architectural future whereby adaptation becomes much more holistic and operates on a very small internal scale.

Acknowledgments. I am very thankful to the valuable experts who consulted on this project in the classroom with supplementary lectures including David Nixon from Altus Associates Architects, Dr. Edward McCullough who gave the guiding premise of the habitat and Dr. Phyllis Nelson from Computer and Electrical Engineering at Cal Poly. Scott Howe from NASA/ JPL was also very helpful for his criticism as well as providing an insightful tour of the Jet Propulsion Laboratory. I would also like to acknowledge the students whose work is shown in this paper including: Sarah Hovsepian, Amanda Schluter, Gregory Ladjimi, Kim Jensen, Zac Noguera, Oleg Mikhailik, Houston Drum and Sergio Marquez.

Development of Information Filtering Systems for Disaster Prevention

Yoshinori Hijikata, Tsutomu Yamanaka, Yuya Tanaka, and Shogo Nishida

Graduate School of Engineering Science, Osaka University,
1-3 Machikaneyama, Toyonaka, Osaka, Japan
{hijikata,nishida}@sys.es.osaka-u.ac.jp
{yamanaka,tanaka}@nishilab.sys.es.osaka-u.ac.jp

Abstract. Text data with spatio-temporal information are becoming common with the popularization of mobile phones with a GPS function and microblog services like Twitter. This study proposes a system supporting operators in a disaster prevention center who control an area in real-world. Our system has three functions: (i) automatic classification that classifies messages into a fixed category, (ii) clustering that aggregates similar messages and (iii) burst detection that detects an event in which messages are arising in high frequency. We asked 120 people to send text data with spatio-temporal information by cell phones in the Osaka Expo Memorial Park. We evaluated our system using the above data.

Keywords: text summarization, spatio-temporal information, clustering, burst detection.

1 Introduction

In recent years, map applications that are available on the web are becoming popular. In addition, sending text messages with GPS information is becoming popular because of the emergences of GPS mail service and the software of microblog for cell phones. Due to this popularization, text data with spatio-temporal information is increasing. In the future, if we can obtain huge amount of text data with spatio-temporal information, we will be able to use the data for a situation assessment when a disaster happens. In this work, we develop a support system for a situation assessment using text data with spatio-temporal information (After here, we describe a unit of this text data as "message").

It is important that how well municipal employees (After here, we describe people who make a situation assessment with this system as "operator") assess a situation of controlled area in a short time when we develop a system. A common method for visualizing the above messages is relating a message to an icon, and mapping the icon on a map based on latitude and longitude. However, we cannot assess a situation with a glance at the map by merely displaying icons of messages.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 318–327, 2011.

There are several types of information sent from unspecified number of users on-site. For example, they are information about fire, materials, traffic and so on. If we overlay common icons regardless of the types of information, we cannot assess expanses and concentration of each problem. Also, it is assumed that a lot of information is provided from the area where huge problems happen. For example, if a riverbank failure happens, a lot of information about it is provided. In a situation assessment, operators might want to distinguish whether each message reports the same thing or not compared with the previous messages. Lastly, if operators try to assess what kind of changes have happened (or is happening), they must assess a lot of messages that have been sent at different times. This seems to be a burden to them.

In this work, we design and develop a support system for a situation assessment to solve these problems. This system has three functions. The first function is classifying messages into predetermined categories automatically using their contents. The second function is clustering similar messages using spatial information and the above category information. And the last function is detecting events that happened in the past or is happening at the moment using temporal information of messages and the above cluster information.

The rest of this paper is organized as follows. In Section 2, we present an overall architecture of our system. We show methods for realizing the three functions of our system in Section 3, and we confirm that our system is effective to assess a situation through experiments in Section 4. In Section 5, we introduce related works, and we conclude this study and show directions of our future work in Section 6.

2 System Design

2.1 Overall Architecture

Figure 1 shows the overall architecture of our system. Our system is a server-client system. We assume that people use handheld devices to send messages, a server receives messages, and the system users use a PC with web browser to assess a situation. We implement the system as a web application. We use Google Maps as a map to display information, and implement the system using Google Maps API [5]. In addition, we use Ajax to present information smoothly. By using Ajax, our system can display necessary information without page transitions when it extracts the information from the database in the server. People in real world send messages to the server using handheld devices. Message bodies, time information and location information (latitude and longitude) are stored in the database by the server. We can get sending time from header in mail transmission protocol. Location information is embedded in bodies in a fixed format if we use a GPS mail service. Each time the server receives messages, it classifies the messages into predetermined categories automatically, and adds the categories to the above database. Our system also conducts clustering and burst detection at constant time interval or in response to operators' requests, and

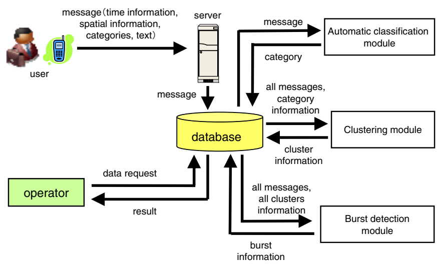


Fig. 1. An overall view of our system

register the results in the database. Operators conduct a situation assessment using a web browser.

2.2 A Method for Utilization of Our System

Once operators invoke the system on a web browser, a map of the controlled area is displayed and messages are mapped on the map in the form of icons. They can assess where and what types of events are happening seeing the distribution of icon colors because the icons are color-coded according to predetermined category. They can view a balloon that includes a text message and time of occurrence clicking each icon. At the right of the map, received messages are displayed in temporal sequence. Therefore operators can check what messages were sent in temporal sequence.

Fig. 2. A screenshot of the proposed system (Individual message view)

When operators click the “content” tab at the bottom left, predetermined categories are displayed in the form of checkbox as shown in Figure 2. By ticking the checkbox, they can restrict messages to be displayed to specified categories. When they click the “period” tab, an input form of period is displayed as shown in Figure 3. By specifying the period, they can restrict messages to be displayed to the specified period. They can also restrict messages to be displayed to a combination of specified categories and period.

By clicking the “normal”, “cluster”, “burst” radio buttons at the bottom right, they can switch display modes of messages. In the “normal” mode, our system assigns an icon to a message, and maps the icon. In the “cluster” mode, our system displays the result of clustering messages. We show a screenshot of

Fig. 3. A screenshot of the proposed system (clustering view)

Fig. 4. A screenshot of the proposed system (burst view)

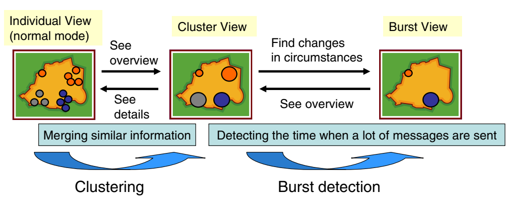


Fig. 5. Interactions in visualization display

the "cluster" mode in Figure 3. Our system assigns an icon to a cluster in the "cluster" mode. The cluster size is represented by the icon size. An icon color of a cluster is a color of the most frequent category in the cluster. When operators click an icon, the size of the icon is displayed in a balloon, and messages in the cluster are displayed at the right message box in temporal sequence.

In the "burst" mode, our system determines whether or not a burst has occurred in each cluster. If a burst has occurred in a cluster, it displays the icon of the cluster. Operators can change the period of the burst detection using the "burst setting" tab at the bottom left. In Figure 4, we show a screenshot of the "burst" mode. We show bursts that occurred from 9:31 to 10:01 in this example. There are six clusters in which bursts have occurred in this example. By clicking an icon, the cluster size and time of the burst occurrence are displayed. A burst means the state that messages are being received more closely than usual in a cluster.

Interaction between users and the system in this study is summarized as Figure 5. In order to assess what is happening in the whole area generally, the "cluster" mode is suitable because it reduces the number of icons summarizing similar messages. If operators find an area with some problems and want to assess its detail, they see icons of respective messages changing the mode into the "normal" mode and confirm the distribution of them. On the contrary, if operators want to check a situation of the whole controlled area while they confirm the distribution of respective messages, they check a whole situation changing the mode into the "cluster" mode. If operators do not have much time to assess the whole situation and want to check whether or not unusual events happen at least, the "burst" mode is suitable because it only informs operators of apparent changes in the whole area. In this way, operators can use our system with changing the modes according to how they want to assess the controlled area.

3 A Method for Supporting a Situation Assessment

In this section, we describe methods for realizing the three functions of our system described in the previous section.

3.1 Automatic Classification of Messages

In this work, a classifier is created by machine learning for automatic classification of messages. We adopt SVM (Support Vector Machine) [6] as a machine learning algorithm. We describe operators' viewpoints as "categories". The categories are previously given by operators. The classifier is created in each category, and determines whether or not each message belongs to the category. Therefore, messages can be classified into several categories. We extract words by conducting morphological analysis to text data. Furthermore, we extract content words (in this work, noun, verb and adjective) from the extracted words. We evaluate mutual information between each word and each category, and make the words into vector elements in descending order of it in each category .

3.2 Clustering

In this work, we conduct clustering using categories which is determined by automatic classification described in the previous section and spatial information which messages have. We adopt Newman's clustering method [7]. This method is a type of agglomerative clustering, and stops cluster merging according to *modularity*. It is often used for finding dense connections between nodes in social networks. We regard a message as a node, and conduct clustering based on weight of edges between the nodes. The weight of edges is calculated from both distances based on sending location and similarities between the messages and the categories.

3.3 Burst Detection

We use a method of Kleinberg's burst detection [8] to detect changes of a situation. We describe a state that time interval of events(messages) is more closely than a normal state as the burst state. In our system, we formulate maximum likelihood sequence from a model consisting of the burst state and a normal state. In addition, we define the degree of the burst that is a measure of the dimension (size) of the burst. And, we detect messages when the degree of the burst peaks, and define the time as the burst time.

4 Experiments

4.1 Data Collection

To assess our system, a huge amount of text data with spatio-temporal information is needed. However, at present, not many people send a surrounding situation with cell phones in certain areas. Therefore, we conduct the experiment in the Osaka Expo Memorial Park, and have test subjects send information of surrounding situation with cell phones to collect the data. In particular, we have test subjects send mails with GPS information. We tell them that any contents are acceptable if the mails are useful for other visitors. The date was from 9:30 to 16:30 on November 16th, 2008, and the number of test subjects was 120. That day was admission free day, and many events such as radio broadcast and free market took place. As a result, we got a total of 2,034 mails. In this work, we set six categories "facility", "play", "sightseeing", "dining", "people", and "event" from the perspective of park administrators.

4.2 Experiments for a Situation Assessment

We conduct experiments to examine whether or not our system is effective for a situation assessment.

Preparations for the Experiments. We apply the data acquired in the Osaka Expo Memorial Park to our system, and examine whether people can assess a situation at the time. What we want to examine through the experiments are as follows.

1. Whether or not displaying the result of automatic classification makes it easier for operators to assess a situation.
2. Whether or not using the "clustering" mode makes it easier for operators to assess a situation.
3. Whether or not using the "burst" mode allows operators to grasp alteration of a situation from moment to moment.

We conduct the experiments to examine the above three questions using four systems. System 1 does not have all of the three functions, and only displays messages in the form of icons . System 2 changes colors of the icons according

Experiments of a situation assessment Name( )
System number( )

Question 1.
Show appropriate spots for following contents on the map. If you write answers to the first question, write "①" on the map, and if you write answers to the fifth question, write "⑤" on the map. In addition, if you need to write multiple answers, write the answers in the form of, for example, "③-1", "③-2".

1. There are two spots where people can see both of autumn leaves and a waterfall.
Write the spots respectively. (10 points)
2. There is a spot where about 30 people take pictures of autumn leaves.
Write the spot. (10 points)
3. There are some playgrounds for children.
Write the spots respectively, and answer what children are doing. (20 points)
4. There is a spot where a drain ditch become clogged and water overflow.
Write the spot. (10 points)
5. There are four main spots about food.
Write the spots respectively, and answer each content of them. (20 points)
6. There are two spots where a music live is performed.
Write the spots respectively. (10 points)
7. There is a spot where a bingo event is held.
Write the spot. (10 points)
8. There are some spots where a photo session is done.
Write the spots. (10 points)

Question 2.
By checking several mails' sent times, you can find tendencies when and what people is doing. Answer the tendencies as many as possible.

Fig. 6. Questions that confirm whether test subjects could assess a situation

to categories which is determined by automatic classification. System 3 has the automatic classification function and the clustering function, and can display clusters grouped similar messages together. System 4 has all the functions. It can display only clusters in which the system detected bursts. The number of test subjects is 28, and we assign 7 test subjects to each system. We ask them to answer questions (in Figure 6) that confirm whether they could assess a situation.

The questions consist of Q1 and Q2, and Q1 consists of eight small questions. Q1 consists of six small questions that ask places, and two small questions that ask both of places and contents. Q2 is a question that asks changes of a situation, and we have test subjects answer tendencies when and what people is doing as much as possible. Q1 is 100-point scale. The small questions that ask places are scored out of 10 points, and the small questions that ask places and contents are scored out of 20 points. Q2 is 10-point scale. The time limit for Q1 is 25 minutes and that for Q2 is 10 minutes.

Results of the Experiments. We calculate the average of the results of 7 test subjects excepting the best result and the worst result in each system. Table 1 shows the results of Q1, and Table 2 shows the results of Q2. In Q1, System 3 achieves the highest average score, and followed by System 2 and System 3. System 1 obtains the lowest average score in Q1. In Q2, System 4 achieves the highest average score, and followed by System 3, System 1 and System 2.

Firstly, we consider the results of Q1. System 2 that displays the categories achieves a better result compared to System 1 that only maps messages on the map. System 1 comes to a worse result compared to System 3 and System 4.

Table 1. The results of Q1

	Test subject number							The average of 7 test subjects	The average of 5 test subjects except maximum and minimum
	1	2	3	4	5	6	7		
System 1	38	88	48	31	58	43	10	45.1	43.6
System 2	65	50	70	78	75	78	50	66.6	67.6
System 3	68	66	74	27	81	74	74	66.3	71.2
System 4	58	76	70	60	41	83	50	62.6	62.8

Table 2. The results of Q2

	Test subject number							The average of 7 test subjects	The average of 5 test subjects except maximum and minimum
	1	2	3	4	5	6	7		
System 1	2	4	7	5	2	4	9	4.7	4.4
System 2	4	4	3	0	3	5	6	3.6	3.8
System 3	8	5	7	3	5	4	4	5.1	5.0
System 4	10	2	8	9	5	5	6	6.4	6.6

This shows that automatic classification is effective for a situation assessment. In addition, System 3 achieves a better result compared to System 2. However, we must judge the helpfulness of clustering carefully because the difference between System 2 and System 3 is not large. And, System 4, which can use all the functions, comes to a worse result compared to System 2 and System 3. We should also consider this fact carefully.

There is small dispersion among performances of test subjects using System 2. However, there is large dispersion among performances of test subjects using System 3 and System 4. Test subjects using System 3 and System 4 used the clustering function positively. In the clustering function, we enable operators to set parameters in the form of multiple choice. The parameters determine cluster cohesion and the ratio between spatial weight and categorical weight. Some test subjects used the clustering function without changing the parameters, others used the function with changing the parameters by trial and error. It appears that dispersion among performances occurred because of the latter test subjects. In System 4, some test subjects used the burst detection function, but it did not work well because Q1 was not a question which detects changes of a situation. This is also considered to be the cause of bad performances of test subjects using System 4.

Secondly, we consider the results of Q2. Because System 4 achieves the best result, test subjects can detect changes of a situation using the burst detection function. Also, System 3 achieves a better result compared to System 1, but System 2 comes to a worse result compared to System 1. We find a tendency that test subjects mainly check the places where there are a lot of messages and confirm changes of the places over time. In other words, it seems that they check a stream of people. Therefore, the result of automatic classification becomes less significant, and as a result, it appears that the difference between System 1 and System 2 is subtle.

We summarize the results of Q1 and Q2 as follows. We showed the effectiveness of automatic classification because System 2 was more suitable for the situation assessment than System 1. In addition, we showed the effectiveness of clustering because System 3 was more suitable for the situation assessment than System 1 and System 2. However, we found a problem that people wasted time and effort for parameter settings. System 4 achieved the best result for the question that detected changes of a situation. We could confirm the effectiveness of burst detection from this. However, some test subjects tried out the burst detection function despite a case in which they need not confirm changes of a situation. This suggests that increasing functions lead to a possibility of wasting users' time and effort.

5 Related Work

Some researchers conduct studies that summarize text data with spatio-temporal information and visualize it. Kurashima *et al.* [1]conduct their study using blogs data. They extract temporal information from their submission time, and extract geographical representation from their text data. They display co-occurrence relation between temporal information and text information on a map. Miyamori *et al.* [2] conduct clustering monthly or weekly data using spatial information, and analyze spoken texts in the cluster. They display the result on a map in the form of semi-transparent circle.

Moriya *et al.* [3]conduct their study using texts including address information and temporal information. They propose a system that generates a map that displays regional situation dynamically according to the texts. Fujisaka *et al.* [4]conduct their study using data of Twitter[1]. They acquire a user ID, temporal information and spatial information using Twitter API. They find regional social phenomena analyzing mobility patterns of several users.

Unlike these studies, our study proposes a system that can classify texts automatically and conduct clustering based on contents and spatial information. In addition, we conduct a burst detection of messages. Our system can assess a situation from various perspective, and it is different from the above methods or systems.

6 Conclusion and Future Work

In this paper, we proposed a system supporting a situation assessment in a disaster using a huge amount of text data with spatio-temporal information. Our system provided three functions on the premise of overlaying messages on a map: automatic classification, clustering, and burst detection. We conducted the experiments, and showed that using the above three functions made it easy for people to assess a situation. However, the clustering function has a problem that parameter settings could waste users' time and effort, and the burst detection function was used even when it was not necessary by some test subjects. In our future work, we plan to support the parameter settings in the clustering function, and restrict the functions according to users' tasks .

[1] Twitter: http://twitter.com/

Acknowledgment

A part of this work was supported by Secom Science and Technology Foundation.

References

1. Kurashima, T., Tezuka, T., Tanaka, K.: Blog map of experiences: Extracting and geographically mapping visitor experiences from urban blogs. In: Ngu, A.H.H., Kitsuregawa, M., Neuhold, E.J., Chung, J.-Y., Sheng, Q.Z. (eds.) WISE 2005. LNCS, vol. 3806, pp. 496–503. Springer, Heidelberg (2005)
2. Miyamori, H., Minakuchi, M., Kawai, Y., Zettsu, K., Kidawara, Y.: Study on Aggregation and Browsing of Ubiquitous City Information using Atomospheric Metaphor. IEICE SIG. Notes of WI2 (14), 101–102 (2009)
3. Moriya, K., Sasaki, S. and Kiyoki, Y.: A Dynamic Creation Method of Environmental Situation Maps Using Text Data of Regional Information. Data Engineering and Information Management (2009)
4. Fujisaka, T., Lee, R., Sumiya, K.: Discovery of User Behavior Patterns from Geotagged Micro-blogs. In: The 4th International Conference on Ubiquitous Information Management and Communication (2010)
5. Google maps API, http://code.google.com/intl/ja/apis/maps/
6. Cortes, C., Vapnik, V.: Support Vector Networks. Machine Learning 20(3), 273–297 (1995)
7. Newman, M.E.J.: Detecting community structure in networks. The European Physical Journal B - Condensed Matter and Complex Systems 38(2), 321–330 (2004)
8. Kleinberg, J.: Bursty and Hierarchical Structure in Streams. In: Proc. of the 8th ACM Special Interest Group on Knowledge Discovery and Data Mining (ACM SIGKDD), pp. 91–101 (2002)

Spatial Design, Designers and Users: Exploring the Meaning of Multi-party Service Cognition

Tom Hope[1], Mizuki Oka[2], Yasuhiro Hashimoto[2], and Myeong-Hee Lee[3]

[1] Tokyo Institute of Technology, 2-12-1-E3-3
Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[2] University of Tokyo, Bunkyo-ku, Tokyo, Japan
[3] Design Team Matt, Tokyo, Japan
tomhope@ipo.titech.ac.jp, lee@mattoct.jp,
mizuki@cks.u-tokyo.ac.jp,
hy@sys.t.u-tokyo.ac.jp

Abstract. This paper examines service cognition from the perspective of the design of public or semi-public spaces. Results are presented from a workshop with students using an approach and system we developed to engage users with the design of their communal space. The paper ends with a brief discussion of these results, which suggest that multi-party service cognition has features that must be addressed in the design of systems, particularly in the context of spatial design.

Keywords: service cognition, design, architecture, workshops, universities.

1 Introduction

Many changes happened in engineering and industry during the end of the 20th century, with the greatest of these continuing globally to the present day. Broadly speaking, industrialized nations moved from heavy manufacturing as a base to knowledge economies [1], coinciding with, and supported by, ever-increasing computing power. While developing economies took up the role of being producers of basic goods in their main industries, the 'post-industrial' economies focused on developing their service sectors [2]. With the rise of 'manufactured risk' [3] comes an urgent need to understand and improve the management and control of highly sophisticated technical systems [4]. In concert with this area of research, studies of the cognition of those involved in service in other areas, and of cognition in organizations more generally have continued. "Service cognition" as a research field is potentially very broad and can therefore denote the cognition of individuals providing services in many varied fields. In this paper we emphasize the other part of service systems—the users of services—and particularly focus on the interrelated nature of providers and consumers/users. Additionally, while research has been done on various systems and settings, there is still a need to more fully understand service cognition in the realm of architecture and design. This paper consequently aims to partly fulfill this need. The paper is organized as follows. In the rest of this section we briefly explore the

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 328–335, 2011.

relationship between service cognition and spatial design. In section 2 we introduce a system we have developed to aid the participatory design of semi-public spaces. In section 3 we present some results and conclude with a discussion in section 4.

1.1 Spatial Design as Service

Although not a new development in theory, in practice the commercial desire to create spaces that can be used for multiple purposes, or "flexible spaces", has developed into a prominent aspect of architecture only in recent years [5]. This can take the form of movable walls and furniture, which allow the use of the space to be changed according to different needs of clients, and may even enable changes to be made within the course of a week or day. Architects are usually involved only in the initial construction of buildings, leaving decisions about use to be done by owners. Hence, with the ability to transform spaces after their construction, if buildings are to be used by people other than themselves, the managers and administrators of those spaces take on an additional role of designers—or, rather, 're-designers'. It is in this sense that the management of a space and reworking of layout is clearly a service. If the space is public or semi-public, such as a gallery, university campus or office, the providers and managers of the space are part of a complex system of provision and use.

1.2 Cognition, Service and Participatory Design

We can note that while traditionally architects and designers of spaces may be separated from the users—the architects and interior designers working prior to users being in the space—administrators of buildings necessarily need to engage with users at some level. However, their evaluations of spatial use might typically take the form of surveys [6]. While this process may be iterative, it does not involve the users directly with the re-designing of the space and may struggle to fully manage to comprehend the multiple layers of spatial use [7]. A participatory design [8] approach may be a way to more fully involved users in the provision of appropriate spatial design. We assert that involving the users in the (re-)designing of the spaces they use will be a useful way to create functional, flexible spaces that 'fit' with changing needs. However, the potential mis-match of cognition between professionals and non-professionals and between users and other users could create difficulties. The appropriate inclusion of digital technologies in this process may benefit the administrators, users, and professional designers.

Studies that use mobile or pervasive technologies to understand spatial use are multiple [9, 10]. Unfortunately, while the data produced is useful and may be generalised, it is difficult to gain access to the actual beliefs or desires of users at the time of use. It should also be recognised that people involved in the service system—designers, administrators and users—will be unlikely to have matching cognition. Indeed, the differences in the beliefs or assumptions about use of original architects and eventual users may be fundamental to why spaces are unused or under-used. In a multi-party process of redesigning space, a system is needed that allows all participants to work together and to somehow 'see' each others' cognition.

In the rest of this paper we therefore do not intend to present a model of service cognition in spatial design, but instead hope to illuminate one way of determining this cognition. We demonstrate a system where members—users or designers of space—can work together in a participatory framework to create new ways of use.

2 The Pingpong Project and Platform

Participants involved in services, whether on the receiving end—in this case the users of a space—or the other side, as designers, architects or administrators, necessarily plan and evaluate at different levels. It has been shown that visualisation, whether in the form of sketches [11] or more elaborate diagrams, is a significant part of design [12]. This visualisation may be considered to be a way to 'externalise' cognition, a practical cognitive map which can be utilised by others in the design process. Our project, known as 'pingpong'[1] aims to structure design processes using multi-disciplinary research related to computer science, and technologies such as web-mining, natural language processing and semantic web. We have developed a platform and set of tools to facilitate this aim, and have to date explored its use in several design workshops held at different university campuses.

2.1 Pingmap Application

For the users, the system consists of a mobile application installed on mobile phones and a web-based browser visualization. The platform itself is the application, a human behavior extraction engine and a visualization map. Upon launching the application, users see a start screen and then a screen with a map of the university campus (or space that will be studied), seen in Fig.1 (a and b).

The user touches the screen to locate themselves on the map, after which an input box is displayed where they insert some text (Fig. 1c). The text is then uploaded to the Web using Twitter, the micro-blogging service[2]. Because we used Twitter as an

(a) start screen (b) campus map (c) text input

Fig. 1. The Pingmap application

[1] http://www.pingpong.ne.jp/
[2] http://twitter.com

underpinning infrastructure to collect the data, the number of characters is limited to 140. This data is collected along with user's account name, posted text, a hash-tag, X-axis of the map, Y-axis of the map, and the floor number of the space. As an example, if a user with an account name of miz oka posts a text "Coffee now", the location information of the map (X-axis, Y-axis, and the floor number) along with the hash-tag specified for the workshop are automatically attached and will produce the following post:

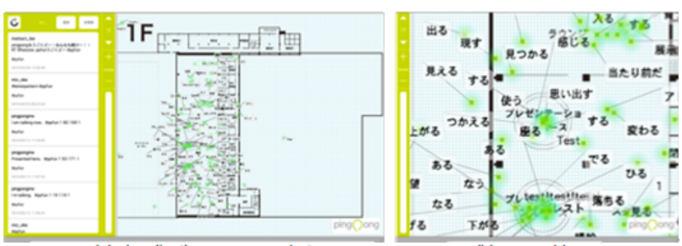
Miz_oka, coffee now, #ppfun, 283, 142, 5. 2010-02-15 10:35:33 JST

2.2 Pingpong Engine

The pingpong engine, a human behavior extraction engine, analyses each post from the pingmap application and extracts human actions. For analysis, we define 'human behavior' in terms of three linguistic elements (subject, verb and object). Each user's posted text (referred to as tweet from herein) is collected using the API and natural language processing techniques are applied to extract these three elements as one human action. Specifically for example, from a tweet "I read a book", the engine extracts "I", "read", "a book". Further details of the natural language processing applied and specifically the methods of dealing with ungrammatical and fragmented text in tweets can be found in [13].

2.3 Pingpong Map

The pingpong map visualizes the collected data in a web-based browser. The visualization is provided in the form of a map in which the verb is extracted from each post and plotted using its location information. This visualization is intended to enable collaboration between users, provided so as to facilitate the sharing of information among workshop participants. Fig. 2 shows a snapshot of the pingpong Map developed for a workshop described below. The left part of the screen shows the tweets in a timeline, but the right side is a map of the space showing only the verbs plotted in the location where the tweets were posted. The reason for showing only verbs is to allow the user to grasp the general content of the tweet and location without needing to read the entire text. Verbs rather than nouns allow the grasping of content in sentences [14]. Clicking on a verb on the map prompts the appearance of its original tweet in the timeline, and clicking on the timeline tweets highlight them in the map for easy reference.



(a) visualization map snapshot (b) zoomed-in

Fig. 2. Screenshot of the Pingpong Map

2.4 Workshop Study

To evaluate the tools in a real-world system, a series of workshops were conducted in university campuses. Here we report on one, held to examine the use of a campus of a university in the north of Japan.

The workshop was conducted with a total of 15 people consisting of 12 undergraduate and graduate students from the university under study and three students from other universities. Students were recruited via an introductory lecture we gave at the university, the mailing lists of the staff and the students, and advertising it on our project website as well as the website of the public relations department of the university. Most of the participants majored in either computer science or information science. The participants were divided into three groups with each group consisting of five students. We also made the set of tools developed for the workshop available for the people who have internal access to the university local network to access, allowing anyone to download the software. This was advertised a month in advance and posted on our blogs as well as the university's website.

The workshop consisted of six phases, namely observation, analysis, idea generation, prototyping, implementation and use. These phases were held on different days, depicted in Fig. 3.

Fig. 3. Day 1: data collection, Day 2: analysis, Day 3: idea generation, Day 4 implementation

On the first day, students used the pingpong application on mobile phones walking around the building. They were told to post tweets freely. This data was collected by the system and used on the second day where workshop participants were provided with the pingpong map visualization tool. Using the tool, the students attempted to understand the current and desired use of the campus space. On the following two days, ideas about new ways of using the space were generated based on the earlier analysis, and finally prototyped to see if they were workable. The workshop participants presented their ideas and prototypes to each other on this last day. Throughout the workshop, hand-held video was captured of student interactions with the system, along with separate audio data during the second and third days of group discussion in order to give observational data for analysis.

3 Results

The workshop proved to be an effective way to see use of the system in a real-world setting. Firstly, the video of the students while collecting data (tweeting) showed that data on spatial usage could be collected in a short time with minimum instruction. It was initially noticeable that the tweets produced by students around the space could apparently be categorized into whether they were about a static observation about the physical space, or whether students—most of whom were regular users of the campus—noticed some change in the space since a previous visit (described more fully in [14]. This is significant as it is information that could not be gained easily by an external designer, but that could prompt users to think of a possible way to redesign the space.

Fig. 4. Students preparing their implementation

The second day was initially frustrating for some of the students, as the pingpong map did not function according to how they desired it to. Initial interest was on the data that they had produced themselves and which would have been the topic of their discussion and design. But the way the tweets had been processed by the system made it very difficult to simply find one's own posts without seeing those of other people. In effect the system pushed workshop participants toward comprehending other users' experience and desire for the space. The freedom that students had been given about their posting of tweets with the system also highlighted the necessity to recognize the different levels of spatiality, namely that in addition to physical environmental observations, social interactions between students led to posting – much as we may expect in non-workshop 'real' settings. This can be seen in Extract 1, which shows two tweets, with the latter one prompted by the sight of the first.

[Extract: Social Interaction Tweets]
Student A
Waving at a person who got off from the elevator. 2010-02-16 13:42:45 JST
Student B
This may be useful for finding a person. 2010-02-16 13:43:27 JST

The beliefs held about the usage of the space were therefore shown to emerge from interactions with others in the same building. While it would not be impossible for individual administrators and designers to gain access to this data, pingong and the workshop format facilitated this easily.

Students used the visualization to generate ideas about redesigning the space to more effectively represent desired use and all were able to decide an implementation to prototype on the final day. These included ways to improve navigation information displayed in the building and to more fully utilize unused areas.

4 Discussion

This study briefly introduced a system to aid in the redesigning of semi-public spaces. Though the study at this stage did not involve professional designers, the workshop participants found the system useful in their development of a new spatial designs. It is envisaged that the logical extension of a system such as pingpong would be the ability to collect data naturalistically, such as that produced in users' daily lives through micro-blogging or other mobile-online behavior. This will be a next stage of our future work. Nevertheless, the workshop itself provided a significant benefit that data could be collected in a short time and the producers of that data were then able to join and collaborate around a representation of their use and desire for use of the space. As such, a workshop format with a system like pingpong may be beneficial to administrators of buildings in the short term at least.

The study illustrates that service cognition in the context of participatory design is complicated by the involvement of multiple parties, each with differing understandings of present and desired use. While the workshop this time included one type of user—the students—the system prompted their recognition of beliefs and desires in the other groups. This would only be emphasized more with the addition of professional architects and administrators. Although the pingpong system has been developed from an emphasis in participatory spatial design, we believe that it may have use in other types of multi-party service, as it can enable, even strongly prompt, participants to visualize and discuss their cognitive maps.

References

1. OECD: The Knowledge-Based Economy, OECD, Paris (1996)
2. McMichael, P.D.: Development and Social Change: A Global Perspective, 4th edn. Sage/Pine Forge, Thousand Oaks, CA (2008)
3. Giddens, A.: Runaway World: How Globalization is Reshaping Our Lives, Profile, London (1999)
4. Kanno, T., Nakata, K., Furuta, K.: A method for conflict detection based on team intention inference. Interact. Comput. 18(4), 747–769 (2006)
5. Kronenburg, R.: Flexible: Architecture that Responds to Change. Laurence King Publishers (2007)
6. CABE: Creating Excellent Buildings: A Guide for Clients. (2011), http://www.cabe.org.uk/publications/creating-excellent-buildings
7. Paay, J., Kjeldskov, J., Howard, S., Dave, B.: Out on the Town: A Socio-Physical Approach to the Design of a Context-Aware Urban Guide. ACM Trans. Comput.-Hum. Interact. 16(2), Article 7 (June 2009)

8. Blomberg, J., Giacomi, J., Mosher, A., Swenton-Wall, P.: Ethnographic field methods and their relation to design. In: Schuler, D., Namioka, A. (eds.) Participatory Design: Principles and Practices, Hillsdale (1993)
9. Randell, C., Muller, H.L.: Low cost indoor positioning system. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, p. 42. Springer, Heidelberg (2001)
10. Lee, S.-W., Mase, K.: Activity and Location Recognition Using Wearable Sensors. IEEE Pervasive Computing 1(3), 24–32 (2002)
11. Gero, J.S., Tversky, B., Purcell, T. (eds.): Visual and Spatial Reasoning in Design II, Key Centre of Design Computing and Cognition, pp. 271–282. University of Sydney, Australia (2001)
12. Won, P.H.: The comparison between visual thinking using computer and conventional media in the concept generation stages of design. Automation in Construction 10(3), 319–325 (2001)
13. Oka, M., Myeong-Hee, L., Hashimoto, Y.: Pingpong: Supporting Collaborative Design Work by Mining Human Behaviors in the Field. In: The Second International Service Innovation Design Conference (2010)
14. Langacker, R.: The Foundations of Cognitive Grammar: Theoretical Prerequisites, vol. 1. Stanford University Press, Stanford (1987)

Toward an Understanding of a Computerized Monitoring System Failure: An Interpretive Approach

Nathan Johnson, Yibai Li, Fengchun Tang, and Saonee Sarker

Entrepreneurship and Information Systems,
Washington State University, Pullman WA 99164
{njohnson76,yibai,fengchun_tang,ssarker}@wsu.edu

Abstract. CMS (Computerized Monitoring Systems) are prevalent in organizations and fulfill an important role to management. Often, new CMS implementations fail. A possible aspect of failures is the perception by employees of being "watched" by CMS. Since all information systems can inherently collect data, they can all monitor employees at some level, even if it's not their primary purpose. Therefore, users might sense they are being watched, triggering panoptic perceptions. This research in progress endeavors to understand the panoptic phenomenon and resulting CMS implementation failure via Actor Network Theory (ANT). Specifically, a published case study on a CMS failure within a medical organization is revisited to investigate organizational dynamics of misaligned interests between the cast of characters. Through ANT, the reader attains a better understanding of the technological and social factors that affected the CMS implementation and the underlying processes that led to its failure. Implications and future research are discussed.

Keywords: panopticon, computer monitoring system, Actor Network Theory, interpretive understanding.

1 Introduction

In 1787, Jeremy Bentham introduced a prison design called the *panopticon*. By design, the prison cells were arranged circularly around a central guard tower so prisoners could be monitored without their knowledge. As described by Bentham, "the persons to be inspected should always feel themselves as if under inspection, at least as standing a great chance of being so"[1].

The idea of panopticon is efficient monitoring, and it is common in organizations. Recent statistics report nearly 80% of organizations utilize some form of monitoring [2]. It's not surprising given workers have been monitored throughout history [3]. Monitoring is employed for a variety of reasons including internal, quality, and performance controls. Tracking employee performance [4] is a typical use which can both reward productive and motivate underperforming employees [5].

Advances in technology have made monitoring more efficient and less obvious. Consequently, electronic tools such as computerized monitoring systems (CMS) have replaced monitoring methods requiring human interaction. While some information systems (IS) are designed for monitoring, most are not intended for this purpose

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 336–345, 2011.

specifically. However, since all IS have inherent abilities to collect data from users at some level, they all are potential monitoring systems. As a result, IS users might sense they are being watched, triggering panoptic perceptions. Timmons [6] referred to this notion as electronic panopticon (EP).

While organizational monitoring, especially CMS, has been examined in the social science literature, little research has investigated why these systems sometimes fail to fully implement or be successful in organizations [7]. A CMS failure is a setback to finances, management credibility, and employee morale [7]. This research uses Actor Network Theory (ANT) to investigate a failed CMS implementation. Guided by Lee's [8] framework of understanding, we provide a subjective overview and a deeper interpretive look at the actor's roles in the failure. Our analysis explores how panoptic perceptions of the CMS played a key role in the implementation failure.

2 Literature Review

A review of literature reveals much work on the panoptic problem. The following is a brief selection. Foucault's "Discipline and Punish: The Birth of the Prison" [3] is one of the earliest studies on panopticon. It has since often been associated with IT aided monitoring. Zuboff [9] coined the term "information panopticon" and positing management control reach past time and space constraints. Lyon [10] used "electronic panopticism" in his systematic description. Botan [11] proposed a panoptic effects model to predict consequences of electronic monitoring in the workplace. There is now general acceptance as stated by Foucault [3] that panopticon is a near inescapable phenomenon in the modern electronic world [6 p.143].

We use CMS as an umbrella term for any IS built to collect data from users. Some IS are created to monitor [12] such as electronic performance monitoring systems [13], security monitoring systems (e.g. CCTV) [14], and call center monitoring systems [6]. In contrast, other systems (e.g.: payment processing systems) are not designed explicitly for monitoring purposes. But because they collect data, they can be used as a monitoring system. Depending on the extent management uses this category of IS, these CMS can be seen as a form of EP. For instance, Kayas [15] looked at the panoptic phenomenon within the scope of ERP systems. Dawson et al. [16] and Kitto [17] studied panoptic effects in online learning environments.

A rich literature exists concerning individuals working in EP environments. Research has shown EP caused by CMS radically increased the stress of an employee, decreased satisfaction, impacted job performance, and influenced the sense of group citizenship and commitment to the employer [13]. Studies have found system feedback and perceptions of fairness moderate these negative effects [18].

3 Methodology

We employ the first two steps of Lee's integrative framework [8] to gain a better understanding of a CMS implementation and failure. Lee's framework proposes three levels of understanding: subjective (perception of phenomenon by the subjects), interpretive (perception of phenomenon by researcher), and positivist (confirmation of researcher's theory). Taken together, a cycle emerges ending with the positivist view validated through a subjective understanding, which initiates a new cycle.

We now present a subjective look at a failed CMS implementation by reporting on a case study from the literature. Following is an interpretive understanding using ANT, looking particularly at actor roles and the panoptic influence.

3.1 Subjective Understanding

Subjective understanding refers to "how the observed human subjects understand themselves and the world around them" [19]. Here, subjective understanding is captured through description of a prior case from the literature [20], a technique not uncommon in the IS discipline [21, 22].

Our case describes a failed CMS implementation at MEDWELL (a hospital pseudonym). MEDWELL's host country underwent major reform in its public healthcare sector. To make healthcare facilities more accountable, the government corporatized all public hospitals and created a Regional Health Authority to act as the final purchaser of health care services [20].

Under pressure to reduce costs and improve efficiency, MEDWELL implemented a CMS. Labelled 'casemix' (CX), it was designed to "track individual patient treatments throughout the hospital on a cost and activity basis" [20]. With information generated CX, management could monitor expenditures, implement changes based on collected data, and normalize medical practice in an effort to reduce costs.

However, deployment of CX was plagued with problems. Instead of seeing it as a way to improve efficiency, clinical staff viewed CX as a system that intruded on their autonomy. As a result, excuses developed to circumvent the use of CX. Worse, some doctors used CX data to lobby management for more expenditure of resources. Although management hoped CX would reduce costs and boost efficiency, clinicians never embraced it. As clinicians gave up using CX, so too did management, resulting in a failed implementation of the CMS [20].

From a subjective standpoint, the CMS failure occurred due to resistance and lack of use: symptoms included negative feelings of EP from the clinicians, misuse of CX data by clinicians, and half-hearted buy-in from management. Normalization Process Theory [23] suggests implementation of organizational process requires cognitive participation and collective action, criteria CX never achieved. Lessened panoptic perceptions of the CMS might have led clinicians to more use of CX. As it happened, clinicians never believed in CX or wanted to participate in its use. Additionally, their use of CX data in opposition to management goals (cost cutting) and management's diminishing use of CX hindered successful collective use of the CMS.

3.2 Interpretive Understanding

To reach an interpretive understanding, researchers determine how humans view themselves and the world around them. In our research, interpretive understanding is achieved through reinterpretation of the subjective view utilizing ANT.

Basic Concepts of ANT. ANT describes society as a collection of human and non-human elements, which form a complex network of actors. Each actor has interests, whether the actor is human or a non-human object. The core concept in ANT is *translation* in which "the identity of actors, the possibility of interaction and the margins of maneuver are negotiated and delimited" [24]. Translation has three stages: *problematization*, *interessement*, and *enrollment*. In problematization, the focal actor - the key actor driving the process and direction of translation - identifies problems and other actors, outlines strategies for solving problems, and establishes the obligatory passage point (OPP). In ANT, OPP refers to a point specified by the focal actor where all relevant actors achieve a shared interest in a post-translation network. During interessement, the focal actor works to convince other actors to join a newly defined network. At enrollment, actors accept the roles negotiated with the focal actor and pass through the OPP. A new, stable actor network emerges as translation completes [25].

The rationale for using ANT to gain an interpretive understanding is broad. ANT stresses the importance of non-human actors and contends they (and their interests) should be treated similarly to human actors. Therefore, ANT is useful given we want to understand the role of the IS in this setting. ANT is also appropriate for analyzing complex phenomenon within an organization; particularly, political, technical, and social networks. In the present study, many factors mutually influenced the implementation of CX, comprising a complex network. Through ANT, we can gain a finer understanding from the management, employee, and technology perspective.

Pre-translation Actor Network. From the ANT perspective, MEDWELL was an actor-network consisting of management and clinical staff (Figure 1). Both actors had specific roles in the pre-translation network. Responsibility of management was administering MEDWELL's day-to-day activities and providing resources to facilitate operations. Clinicians focused on taking care of patients based on their autonomous judgment. Through different roles, they shared the same interest of providing health care.

Translation – problematization. *Setting the OPP* – Corporatization of MEDWELL triggered the translation process. While the interests of clinical staff remained, management's, due to the new mandate to make healthcare more accountable, changed to reducing costs and improving efficiency. This altered focus caused a misalignment of interests among actors, leading to the breakdown of the old network.

Management decided CX CMS was the way to reach the new costing and efficiency goals, citing a lack of the information need to control costs. Since CX "links detailed information on hospital patient treatment and clinical activity with associated costs…" [20], the CMS would enable them to monitor and control costs. As described, "scrutinizing clinical procedures and explicitly linking patient treatment decisions to standard costs make clinical activity visible and susceptible to intervention by management, who can then influence clinical decisions" [20].

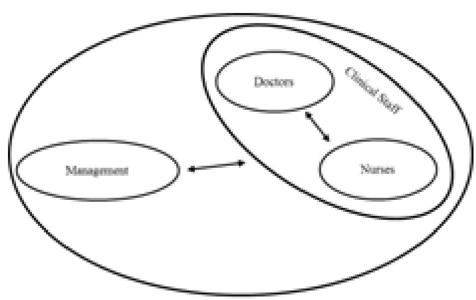


Fig. 1. Old Network (Pre-Translation)

In addition, management thought CX would reduce costs by normalizing doctors' medical practice. As described, "a CX information system is an attempt to influence doctors' behavior towards 'normal' work practices, through comparative application of performance information" [20]. Ideally, this information would engender peer pressure, coaxing doctors to adopt best practices thus reducing costs. A corporate manager reinforced this idea: "There are difficulties overall with actually managing doctors. I believe the only way of managing doctors is to get information through IS that provide them with the sort of reports in which peer pressure will bring some conformance to expenditure" [20]. With this in mind, management (the focal actor) set the OPP for a new actor network: installation and use of CX by all hospital staff.

Identification of relevant actors - Hospital management can be viewed as the focal actor in the network. At MEDWELL, management drove utilization of CX. Management pushed other network actors behind their interest of cost reduction. The clinical staff is another actor in the network. This *punctualized* actor consists of the doctors, nurses, and medical consultants who work within the hospital. Their interests lied in the autonomous delivery of health care to their patients. The CX CMS was a new actor in the network with its own interests. Making inanimate objects actors is an important function within ANT. The question arises, how can non-living objects have interests? ANT posits that non-living objects have interests inscribed upon them. In this case, though it was designed as a cost accounting and tracking system, CX was inscribed by management with the interest of watching over MEDWELL activities. Through the focal actor's inscription of this interest, the CX system acquired its panoptic qualities that the clinical staff would soon sense throughout the hospital.

Translation – Interessement and Enrollment. *Interessement and Enrollment of the CX system* - The enrollment of CX was achieved through the inscription of management's interest into the system. For instance, via its omnipresence throughout MEDWELL, CX could associate costs with each activity in the hospital making them visible and calculable. As described, "The increased visibility...enabled management to focus attention on profit or loss-making areas" [20]. Additionally, CX could compare doctors' practice, giving "a view on clinical practice that highlights variances between the performance of individual doctors or clinical specialties" [20].

Interessement and Enrollment of the clinical staff - Interessement and enrollment of clinicians into the new network (Figure 2) was more difficult than CX given their interests conflicted with management. As an interviewed manager stated about MEDWELL clinicians, "They have no interest in either the financial side of CX or anything else. They're here to work. They're here to operate. They're here to look after the patient." [20].

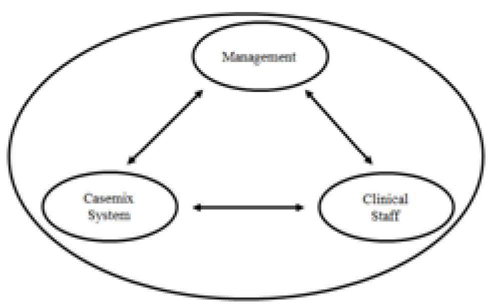


Fig. 2. Ideal State of New Network

Two strategies were utilized to enroll clinical staff into the new network. The first was to exhibit the benefits of CX through a series of demos to clinicians. "Recourse was made to a quality discourse in an attempt to construct the interests of doctors as congruent with those of management. This involved translating the notion of patient care into terms of effectiveness and efficiency" [20].

The second was to recruit senior doctors as clinical managers. Management hoped that some ownership of the CX data by clinical staff would reduce perceptions of being "watched" by managers wanting to control them, thus reducing CX resistance. As described, "The recruitment of senior doctors into managerial roles theoretically meant that control and review would be performed by the same people. Doctors, or at least a 'doctor as manager' hybrid, would influence clinical activity, rather than direct control by senior management". [20]

Failure of the new network. During CX's deployment at MEDWELL, intrinsic conflicts in the new network were never solved. As a result, resistance and betrayal of actors became an inevitable ending. As actors betrayed the new network, an interesting thing happened: the focal actor (management) betrayed themselves and the CX system (Figure 3).

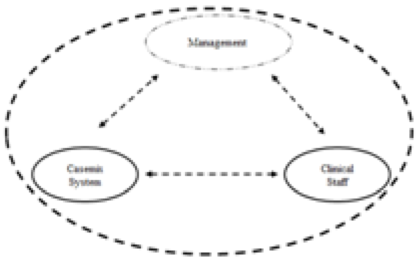


Fig. 3. New Network Disintegrates

Resistance and betrayal of the clinical staff – Two important reasons led to the resistance of clinical staff to CX. First, the panoptic eye of the CX system reached every corner of the clinician domain. No matter the clinical procedure or activity they undertook, no matter where in the hospital it occurred, the ever present CX system was recording their actions as they entered in patient care information. Indeed, CX required clinicians to input information about patient movement, "length of stay, day surgery vs. inpatient surgery, operating theater time management, off-hours laboratory usage" [20]. Compounding the problem, clinicians were unsure how CX data would be used. As related by a corporate manager: "There is a lot of information that is produced, that we're not quite sure, you know, what it means for us… some of my older colleagues felt it was threatening, that you were going to be looking at some of their cases… the fear is that it might be used by somebody against them somewhere down the track… we're not quite sure how that information will processed" [20]. Without a clear management policy as to how information gathered by CX would be used, clinicians were cognitively averse to cooperating with management on use of the system.

Second, although clinical staff members didn't know all the ways the CX data was being used, they were certain about one use of it. CX's constant watch over MEDWELL operations became an affront to the autonomy and best judgment of the clinical staff. As a corporate manager stated, "[Doctors] don't like this monitoring business…they don't like managers and analysts saying 'Hey, why are you doing this, why are you doing that?'" [20].

With CX constantly recording clinical decisions through its data collection, management could choose practices in line with their goals and reject others. Armed with CX data, management began making decisions about clinical practices, intruding on clinicians' professional independence. Clinicians clearly felt the panoptic eye on their backs, eventually leading to betrayal of the new network. As related: "monitoring through the CX IS was open to the circumvention of doctors as they attempted to escape the implications of the gaze of normalizing judgment" [20].

Betrayal of the CX IS - CX had its own interests inscribed by management: that of watching over hospital activities to help reduce costs. Surprisingly, CX betrayed the focal actor (management) after its implementation. Instead of leading to cost savings, CX data was used by clinicians to make a case for more resources. This data use was a reversal of the intentions of management to cut costs! As described, "Even skeptical doctors could recognize the usefulness of the CX IS in arguing for more resources" [20]. As one doctor said in an interview, "I won't say it's not useful, because the clinical manager is able to show that we're grossly under-funded… That's useful. I'll accept that" [20].

Management, in turn, betrayed the CX system. As CX implementation progressed, management never completely supported the CMS to make it work across all hospital departments. The organizational "culture change" never completely happened in the minds of the administration [20]. As a result, management eventually stopped its use of the CX system and returned to traditional budgeting methods!

4 Discussion

As noted above, any new IS roll out will be costly to an organization. Implementation failure can be devastating, both to organization finances and employee morale. As such, investigating these failures is useful. In this particular case, the perception of EP by clinical staff played a significant role in the implementation failure.

Since all IS have capabilities to collect user data, each has the potential to create a panoptic perception. Here, the CX system was designed and implemented to reduce costs by normalizing clinician's routines to best practice. Management failed to realize (or chose to not care) that CX was perceived as an EP by the clinical staff. As a result, they didn't take steps to remove or diffuse the perception. The EP drove resistance to the system in a number of ways, and paved the way for the breakdown of the new actor network that was created when the focal actor implemented the CMS.

Management might have reduced panoptic perceptions of the system by making business rules for using CX data more transparent such as policies to explain what data would be collected, how it would be used, who could view the data, and what procedures would be taken to protect the privacy and identity of users. By making transparent rules, the focal actor might have reduced clinicians' unease with a system that constantly recorded their every clinical activity and brought them more in line with management interests, as suggested by Alder and Ambrose [18].

After examining the failure of CX at MEDWELL, an unanticipated ANT finding emerged in our investigation. Although it did not involve the EP, it does underpin the failure of the CMS implementation. Management failed to get completely behind the CMS to make it work across all hospital departments. As related in the case, a culture shift towards CX never took place among management [20]. This critical lack of follow through led management to eventually stop using the CX system and return to traditional budgeting methods. In effect, the focal actor betrayed the very actor (the CMS) brought into the network to sustain their goal of cost reduction!

This “self-betrayal” event raised an interesting question in our minds about ANT theory in general: can a focal actor betray its own interests within the network and still maintain a viable network? If the focal actor’s goals and stated OPP are the driving force behind the creation of a new network, where the enrollment of other actors is necessary to sustain it, can a network be established and survive when the focal actor does not get behind and support their own goal 100%? We believe the answer is no; it is nearly (if not completely) impossible for a new network to maintain integrity without the complete support of the focal actor. We also believe the interpretation of self-betrayal by the focal actor to be a first in the ANT literature.

5 Limitations and Conclusion

Our primary limitation is that our understanding is derived from a secondary case analysis. The author of the original paper may have biased our understanding through his writing. However, the use of a secondary case can achieve the same rigor level as first-hand investigation. Dilthey [26] contends that the task of interpretation and understanding is “an empathetic grasping, reconstructing, and re-experiencing by one human mind of mental objectifications produced by other human minds” [27]. From this perspective, secondary case analysis and first-hand investigation are both subjective in understanding, which intends to recover the author’s originally intended meaning. The only difference is the text of the understanding. The text of the first-hand investigation is the original case scenario, whereas the text of secondary case analysis is the original author’s writing. To summarize, a secondary case analysis could achieve rigor as long as it recovers the author’s originally intended meaning.

How a CMS engenders a panoptic experience for users and how such perceptions influence users’ performance are still not well understood. The current research is a step towards gaining a better understanding of this phenomenon. It provides some clarity to IS researchers and practitioners on the role of technology in creating panoptic perceptions and the negative impact they can have. In our case, panoptic perceptions played a key role in the failed implementation of the CMS at MEDWELL. Additionally, this study provides suggestions on reducing panoptic perceptions, thus improving the success rate of future implementations. Since implementation failures are costly to organizations in several ways, a better understanding of this phenomenon is both timely and useful.

Future research will take us to Lee’s [8] third leg of insight into the CMS/panopticon phenomenon, the positivist understanding. Investigation will center on how system characteristics influence users’ panoptic perceptions. An experiment will seek to measure subjects using in a simulated CMS [28].

References

1. Bentham, J., Bowring, J.: The works of Jeremy Bentham. In: Tait, W., Simpkin, M. (eds.), Edinburgh, London (1843)
2. D’Urso, S.C.: Who’s watching us at work? Toward a structural-perceptual model of electronic monitoring and surveillance in organizations. Communication Theory 16, 281–303 (2006)

3. Foucault, M.: Discipline and punish: the birth of the prison. Pantheon Books, New York (1977)
4. Stanton, J.M.: Reactions to employee performance monitoring: Framework, review, and research directions. Human Performance 13, 85–113 (2000)
5. Stanton, J.M., Julian, A.L.: The impact of electronic monitoring on quality and quantity of performance. Computers in Human Behavior 18, 85–101 (2002)
6. Timmons, S.: A failed panopticon: surveillance of nursing practice via new technology. New Technology, Work & Employment 18, 143–153 (2003)
7. Alder, G.S.: Ethical issues in electronic performance monitoring: A consideration of deontological and teleological perspectives. Journal of Business Ethics 17, 729–743 (1998)
8. Lee, A.S.: Integrating Positivist and Interpretive Approaches to Organizational Research. Organization Science 2, 342–365 (1991)
9. Zuboff, S.: In the Age of the Smart Machine. Basic Books, New York (1988)
10. Lyon, D.: The electronic eye: the rise of surveillance society. University of Minnesota Press, Minneapolis (1994)
11. Botan, C.: Communication work and electronic surveillance: A model for predicting panoptic effects. Communication Monographs 63, 293–313 (1996)
12. Aiello, J.R.: Computer-Based Work Monitoring - Electronic Surveillance and its Effects. Journal of Applied Social Psychology 23, 499–507 (1993)
13. Bates, R.A., Holton, E.F.: Computerized performance monitoring: A review of human resource issues. Human Resource Management Review 5, 267–288 (1995)
14. Fussey, P.: New Labour and New Surveillance:Theoretical and Political Ramifications of CCTV Implementation in the UK. Surveillance & Society 2, 251–269 (2004)
15. Kayas, O.G.: The panoptic gaze: analysing the interaction between enterprise resource planning technology and organisational culture. International Journal of Information Management 28, 446 (2008)
16. Dawson, S., Burnett, B., McArdle, F.: Watching Learning From Behind Closed Doors: The Impact of Surveillance on Student Online Behaviour. In: ELearn 2005: World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education (2005)
17. Kitto, S.: Translating an Eletronic Panopticon Educational technology and the re-articulation of lecturer-student relations in online learning. Information, Communication & Society 6, 1 (2003)
18. Alder, G.S., Ambrose, M.L.: An examination of the effect of computerized performance monitoring feedback on monitoring fairness, performance, and satisfaction. Organizational Behavior and Human Decision Processes 97, 161–177 (2005)
19. Lee, A.S.: Electronic Mail as a Medium for Rich Communication: An Empirical Investigation Using Hermeneutic Interpretation. MIS Quarterly 18, 143–157 (1994)
20. Doolin, B.: Power and resistance in the implementation of a medical management information system. Information Systems Journal 14, 343–362 (2004)
21. Gallivan, M.J.: Striking a balance between trust and control in a virtual organization: a content analysis of open source software case studies. Information Systems Journal 11, 277 (2001)
22. Gallivan, M.J., Depledge, G.: Trust, control and the role of interorganizational systems in electronic partnerships. Information Systems Journal 13, 159 (2003)
23. May, C., Finch, T.: Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory. Sociology 43, 535–554 (2009)

24. Callon, M.: Some Elements of a Sociology of Translation - Domestication of the Scallops and the Fishermen of St-Brieuc Bay. Sociological Review Monograph, 196–233 (1986)
25. Sarker, S., Sarker, S., Sidorova, A.: Understanding Business Process Change Failure: An Actor-Network Perspective. Journal of Management Information Systems 23, 51–86 (2006)
26. Dilthey, W.: Selected writings. Cambridge University Press, Cambridge (1976)
27. Prasad, A.: The contest over meaning: hermeneutics as an interpretive methodology for understanding texts. Organizational Research Methods 5, 12 (2002)
28. Sia, S.K., Tang, M., Soh, C., Boh, W.F.: Enterprise resource planning (ERP) systems as a technology of power: empowerment or panoptic control? SIGMIS Database 33, 23–37 (2002)

Proposal of BCM Evaluation Method Based on Disaster Scenario Simulation

Ryuhei Kaneko and Yoshio Nakatani

Graduate School of Engineering Science, Ritsumeikan University,
1-1, Noji-Higashi 1, Kusatsu, Shiga 525-0058, Japan
cc005063@ed.ritsumei.ac.jp, nakatani@is.ritsumei.ac.jp

Abstract. Almost 20% of big earthquakes in the world occur in Japan. It becomes impossible for private companies and governments to carry out their business if they suffer from severe damage. In this situation, BCM (Business Continuity Management) become focus of public attention. But it has not become popular enough. This is due to the fact that anti-disaster plan is difficult to be verified because companies are lack of knowledge about what situations may occur in disasters. This research proposes how situation changes in companies by using a disaster progress simulator and actual cases when disasters happen. By using this system, companies can study what situation changes and clarify problems of their BCMs. As a result, the companies can settle on more realistic BCMs and carry out business even in disasters.

Keywords: Business Continuity Management, disaster simulation.

1 Introduction

A lot of earthquakes occur in Japan. It becomes impossible for private companies, social infrastructure providers, such as electric power companies, gas companies, network companies, and road/railway companies, and governments to carry on their operations when they suffer from severe damages in big earthquakes. In this situation, BCM (Business Continuity Management) become the focus of public attention in Japan. BCM is the activity that settles on policies and procedures to achieve business continuance of companies in contingencies such as the natural damage and terrorism. Currently, Japanese companies are growing interest in BCM, especially after the 9.11 terrorism. But it is not popular enough. This is due to the fact that anti-disaster plan is difficult to be verified because companies are lack of knowledge about what situations may occur in disasters.

This research proposes an anti-disaster planning support system based on disaster scenario simulation which simulates what kind of situations may occur about companies when a disaster occurs. Then this research aims to provide the companies with an efficient environment to investigate BCM enough.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 346–353, 2011.

2 BCM

2.1 What Is BCM

This system supports BCM of companies and governments. BCM is realized by BCP (Business Continuity Plan), which is an action plan of BCM. BCP is required to improve on regular according to changing conditions and knowledge accumulation about disasters because BCP should be effective. The PDCA (Plan Do Check Action: PDCA) cycle has to be rigorously executed as a method of improving them continuously.

Figure 1 reveals that companies innovating through BCM can improve the expedition of their re-establishment and rundown of their capacity operating rate [1].

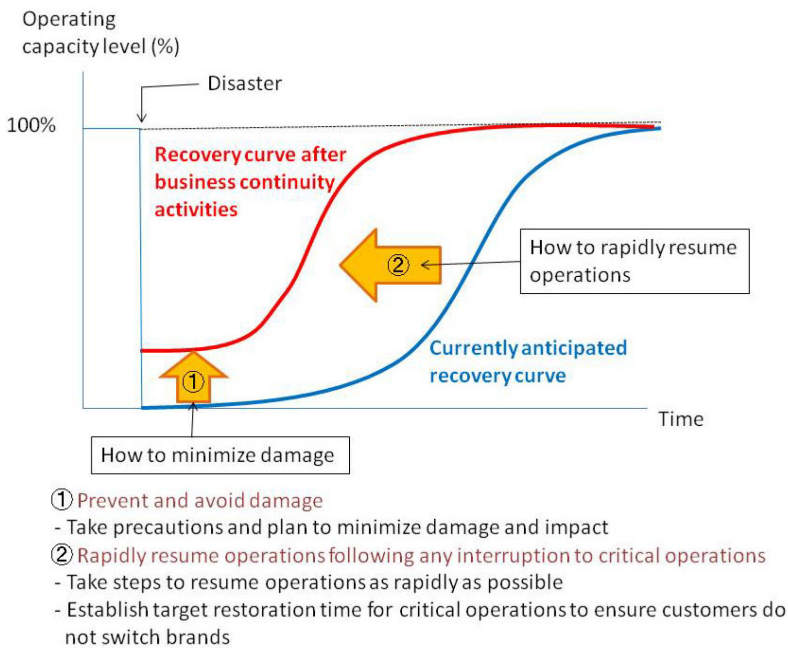


Fig. 1. Effect of BCM [1]

2.2 Actual Conditions Regarding BCM

In Japan the numbers of company realizing necessity of BCMs are increasing. The numbers of companies settling on BCMs are also increasing. Therefore, it can be increase the company having interest in BCP.

However, BCP can be made more effectively by regularly reviewing and improving it. The figure.2 gives the number of Japanese companies that don't have much training about BCPs [2].

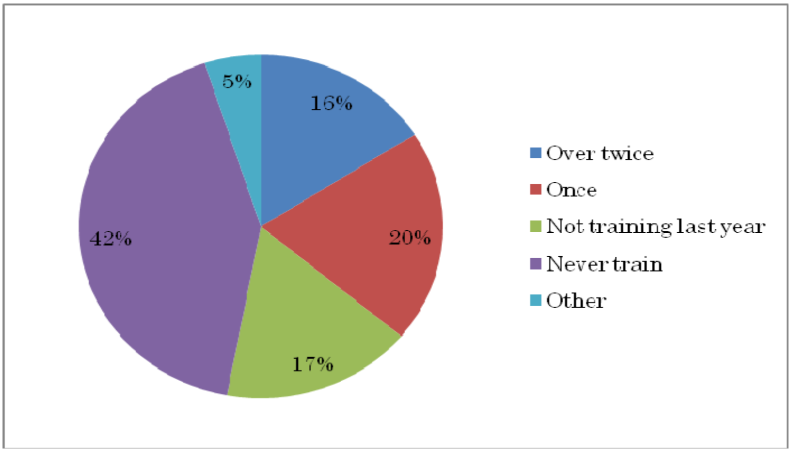


Fig. 2. Graph of BCM education and study

2.3 Related Researches

Research and development of BCM support is currently an attractive and important field and hence the amount of research on it is gradually increasing, although still can be said to be insufficient. An example of the research on supporting BCMs is an earthquake emergency response scenario simulator [3]. It suggests damage situation based on an assuming earthquake scenario. It utilizes a visual imaging database to provide support for making judgments and decisions during the state of emergency after an earthquake. Then a user can plan disaster recovery from damage situation.

2.4 Problems on Supporting BCM

BCMs in Japan include several problems, details on which are provided below.

The number of companies that realizes necessity of BCM and settles on BCM is increasing. However, many companies don't continue to improve and revise BCMs after formulating them. One reason is a problem with difficulty in verification whether they are useful and comprehensive even if BCM is formulated. Especially companies have apprehensions about what situation may break out, even if companies can consider various kinds of countermeasures, based on lack of knowledge.

This apprehension is not reduced enough by previous researches. The computer simulation is effective to call up an image of situation changes in a specific scene, but it is difficult for a company to understand the entire damage that may be caused by a certain level of disaster.

3 Prototype System

3.1 System Summary

This research proposes the system that provides a company with a whole image of possible damage.

This system simulates the development of damages, pointing out what changes may take place in which human resource and equipment of a company, based on general knowledge on causal chain of damages. The user select a part of causal chain by selecting human resource and equipment of his/her company as a initial data, from a list of possible resources that may exist in an average company of a certain area. This causal chain doesn't consider a temporal element, and state transition from damage on a resource to damages on other resources is reasoned. However, the user can assume how long a state does on, because a temporal element is essential in disaster.

System architecture is shown in Figure.3.

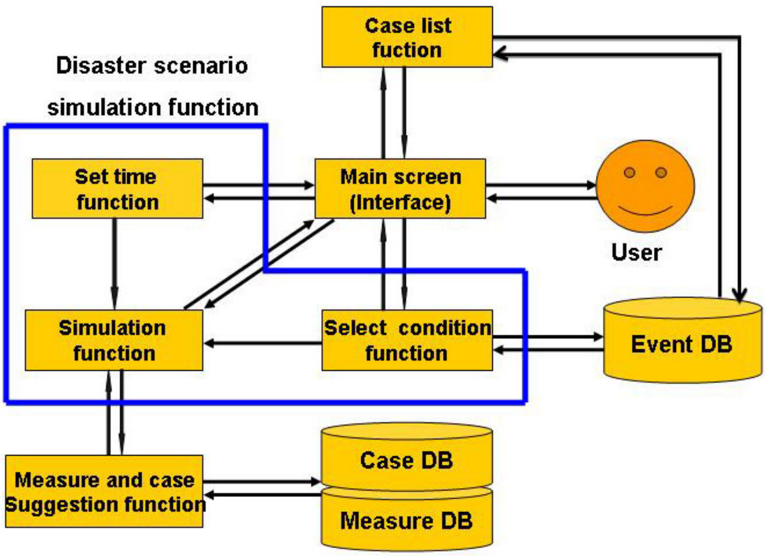


Fig. 3. Image of system architecture

A company can rate the BCM. This system provides the user with functions that change the setting of simulation: "condition selection function" that selects resources from a list of possible resources, "time-setting function" that specifies how long a state goes on, "simulation function" that executes simulation based on specified conditions, and "measure and case suggestion function" that suggests measures and suggests related previous cases.

This system was implemented with Java and PostgreSQL on Windows OS.

3.2 Data Form

This system holds several databases to store the data of events, cases and measures.

1. Event selection database. This database is for which conditions are selected. The data include the basic attributes of a company: building (office/factory) and location (city part/mountainous district/sea surrounding/innings). The assumed hours of simulation is also specified (morning commuting hours/ business hours/ evening commuting hours).

2. Detailed event database. Event names, their IDs and categories are stored. Several events that may occur on certain resources in certain disasters are stored. Resources are described in a level of general categories, such as PC (personal computers) and telecommunications equipment.
3. Case database. Details of the previous cases (places, categories, preceding events, and dates when the cases occurred) are stored in this database. The preceding events are used as conditions that triggered the case.
4. Measure database. Measures to possible events are stored.

3.3 Condition Selection Function

This function can allow setting of the basic condition of a company and assumed hours of simulation. The system selects the appropriate events from the detailed event database based on the selected conditions. Then the system selects the events succeeding the selected events in the database. This selection continues until the appropriate events are not found.

When the system selects the events, the same events are selected every time, because the conditions are the same. The system selected the events randomly among the candidate events.

Figure.4 is the examples of select condition function.

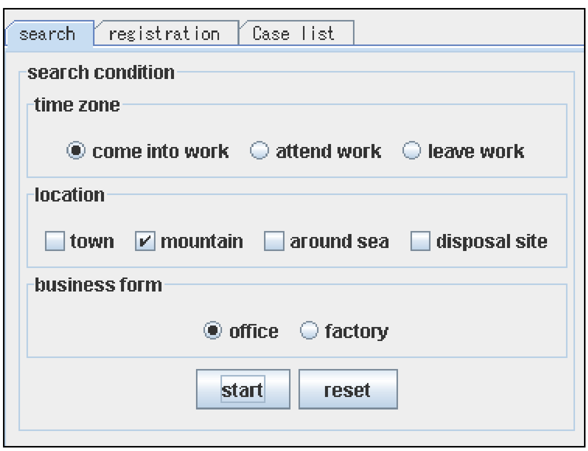


Fig. 4. An example of the condition selection function

3.4 Time-Setting Function

The system holds the data as a common scenario for the development of disaster, but it doesn’t consider temporal element of how long each event goes on. Therefore the user can set the duration to each event.

First, the user selects the event that he/she wants to set duration. Next, the system proposes possible durations that the user can select based on the relationship between the preceding and succeeding events. The possible duration is proposed based on the end time of the preceding event and the start time of the succeeding events. The users can set the duration within the proposed period. Then the system shows the duration in the Gantt-chart on the screen.

Figure.5 is an example of the time-setting function.

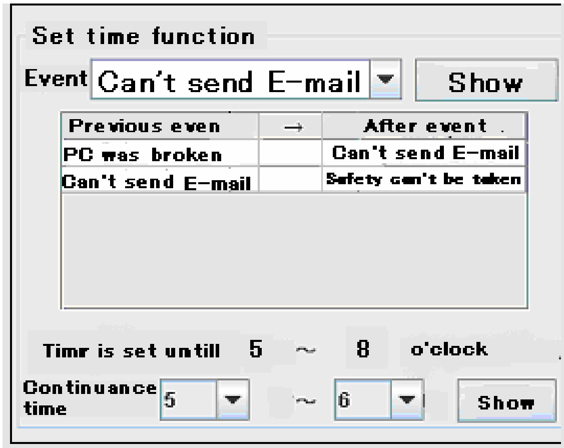


Fig. 5. An example of the time-setting function

3.5 Simulation Function

This function is the main function. This function simulates the development of damages based on the specified conditions and selected events. When the user pushes “simulation button,” simulation starts.

The system follows the event chain in the detailed event database and randomly selects some of the appropriate events, until the succeeding events are not found. For

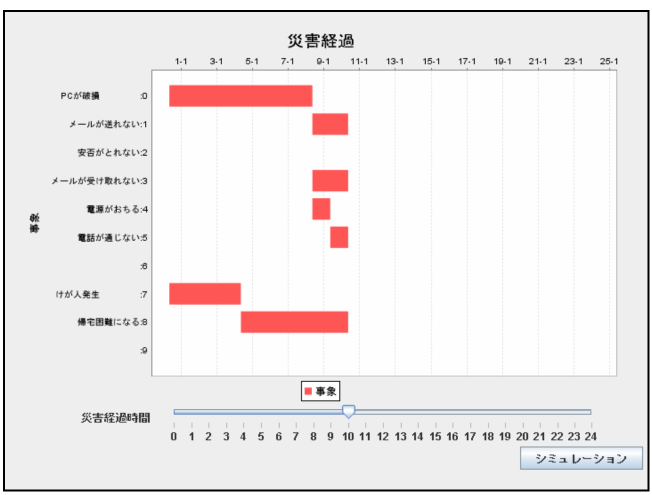


Fig. 6. An example of the simulation function

example, the system reasons “damage of PC falling in the water” based on the transit from “quake” to “falling of PC” and another transit from “quake” to “damage of a water pipe.” The system further reasons “impossible to send e-mails” based on “damage of PC falling in the water.” The result is shown in the Gantt-chart.

Figure.6 is an example of the simulation function.

3.6 Measure and Case Suggestion Function

This function is another main function. The disaster scenario simulation gives the user the damage image but the user doesn’t know how to take countermeasures against possible damages. This function suggests a related previous case and a possible countermeasure. The case shows “what happened at other company when the event occurred” and “what caused the event.” In addition to this, the system shows a possible common countermeasure in the measure database.

This function supports revision and improvement of the BCP.

Figure.7 is an example of the measure and case suggestion function.

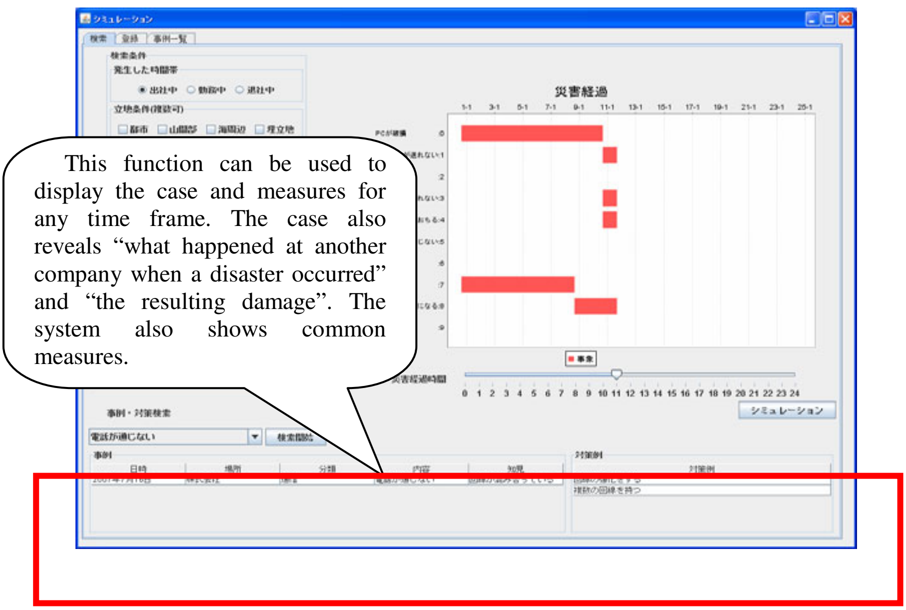


Fig.7. An example of the measure and case suggestion function

4 Result

The systems were evaluated by 2 presidents of a software development company and an insurance company in Japan.

The scenario situation simulator and the measure and case suggestion function were evaluated.

4.1 Evaluation Regarding Scenario Simulation Function

- It is a simple structure, easy to determine what has occurred
- The user can image whole disaster but is difficult to understand the relationship between events.
- Events should be shown by classified into their categories for easiness to understanding.
- It is necessary to show the event dividing into two frames of local risks and common risks.

The scenario simulation function has strictly opinions. In the future, how to show the event is considered.

4.2 Evaluation Regarding Measure and Case Suggestion Function

1. This function realizes continuous revision of the BCP by considering the previous cases and common countermeasures.
2. In accordance with the division of events by categories, the cases and countermeasures should also be classified into categories.

This function is effective but it is necessary to consider how to show the events and cases in accordance with the scenario simulation function. It is necessary to support BCP more effectively. Therefore the system will have a function about rate and support BCP.

5 Conclusion

This research proposes the damage scenario simulation to support revision of BCP in term of what damage may occur and how the damage develops.

The evaluation of the prototype system by the two entrepreneurs shows some good result but some improvement is pointed out. In the next step, improvement will be considered. The system should be evaluated by installed at company. Through actual use, the system has to be improved in the future.

References

1. NEC: Approach to Business Continuity, `http://www.nec.co.jp/csr/en/report2008/continued/index.html` (referred February 3, 2011)
2. NTT-AT: Consider consulting to help implement recovery measures precaution measures "Business Continuity Planning Consulting Services, `https://www.nttat.co.jp/product/bcp/index.html` (referred August 29, 2009)
3. Naganoh, M., Taniai, M., Ikeda, J.: Earthquake Emergency Response Scenario Simulator. Computational Studies A 2, 157–167 (1998)

Design of Communication Field for Leading to Satisfied Understanding: Example of High-Level Radioactive Waste Disposal in Japan

Hiroshi Kimura and Masashi Furukawa

Department of Nuclear Engineering and Management, School of Engineering,
The Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
kimura@nuclear.jp

Abstract. A disposal of high-level radioactive waste produced by nuclear power generation is a pressing issue in Japan. Especially, the process of selecting the siting areas for the HLW geological disposal proceeds with difficulty. One of reasons of this difficulty may be that many people don't have enough knowledge about HLW disposal to evaluate the issue completely. In this research, we introduce one of communication field design in order to lead to satisfied understanding concerning HLW disposal.

1 Introduction

In Japan, nuclear power covers 30% of Japan's electricity generation, and is one of the CO_2-reduction measures. On the other hand, the disposal of high-level radioactive waste (HLW) produced by nuclear power generation is a pressing issue. The Nuclear Waste Management Organization of Japan (NUMO) was organized in October 2002 as an implementing body for disposal of HLW. In December 2002, NUMO started "Open Solicitation," which is one of the processes of selecting the siting area for HLW geological repository. Part of this process includes local governments applying for the survey voluntarily in order to decide the HLW repository. In January 2007, Toyo-town, Kochi-prefecture submitted the application for the first time in Japan. However, it was withdrawn after a change in the local government in April 2007. This illustrated how the process of selecting the siting areas for the HLW geological disposal proceeds with difficulty in Japan.

Many researchers said that one of reasons of this difficulty is people's concerns with HLW in Japan. Many people are anxious about HLW, and oppose hosting the HLW geological repository site. People are generally more fearful of HLW than of the nuclear power plants[1]. Additionally, many people don't have enough knowledge about HLW disposal to evaluate the issue completely[2]. Even if someone is aware of the issue concerning HLW disposal, it is difficult for them to find the information that satisfies their personal concerns regarding HLW issues.

In this research, we tried to design a communication field in order to lead to satisfied understanding concerning HLW disposal.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 354–359, 2011.

2 Design of Communication Field

The form of communication field we designed is a one-day workshop. The workshop is designed for people who are aware of the issue concerning HLW disposal. This workshop consists of four sessions; a lecture session, a question-and-answer session, an opinion formalization session, and a communication session.

In the lecture session, we prepare three lectures for the workshop's participants. The themes of lectures are "Why do we use the nuclear power generation at all?", "Structure and Safety of geological disposal", and "Promoting process of HLW disposal project and Symbiotic system with the siting areas". These themes are decided based on the results of eleven focus group interviews conducted from May to October, 2008. The purpose of this session is to aid the participants to acquire the general information concerning the issue of HLW disposal. Furthermore it allows them to note any questions that come up and any points they seek clarification on during the lectures.

In the question-and-answer session, the participants ask the experts questions about the points which they noted in the lecture session. We set up the three booths relating to each lectures' theme. The participants can select a booth and move among these booths freely. Additionally, we set up an information booth. In the information booth, the participants can be advised by the information advisers on what they should ask, which booths they should go, and so on. The question-and-answer session is the most important in this workshop, because it may lead to the participants' satisfied understanding concerning HLW disposal. So, we prepare to take sufficient time.

In the opinion formalization session, each participant clarifies his or her opinions about HLW disposal project. We prepare an answer sheet on A1 paper for each participant. The top of answer sheet is titled, "What do I think and how will I act regarding HLW disposal issue?", and the participants write down their answers and supporting reasons on the answer sheet. This process allows the participants to clarify their personal thinking and formalize their opinions.

In the communication session, the participants deliver and discuss their answers and opinions concerning the HLW disposal project. First each participant represents their answers and opinions using the answer sheets. After that, all of participants start to discuss each other. We prepare a facilitator to lead this session to proceed smoothly. The experts cannot participate in the discussion because the participants can discuss anything what they think freely. At the same time, the experts can learn what the participants think by listening to their discussion carefully.

3 Operation of the Workshop

This workshop was conducted on June 20th, 2010 in Tokyo, Japan. Twenty citizens participated in this workshop. All of participants are aware of HLW, and almost of them had learned about HLW disposal shortly before the workshop. The experts included a scientist, NUMO's engineers, NUMO's spokesmen, officers of Agency for Natural Resources and Energy. The information advisors were handled by the authors. Additionally we commissioned the professional of facilitation for the workshop's facilitator.

Table 1 shows a schedule of the workshop. Organizer of this workshop was the leader of a NPO concerning environmental issues.

In the lecture session, we prepared three suitable speakers of each lecture topic. The speaker of first lecture, "Why do we use the nuclear power generation at all?", was a officer of Agency for Natural Resources and Energy. The speakers of second and third lectures were a engineer and a spokesman of NUMO.

Table 1. Schedule of the workshop

Time	Session	Details
9:30-10:00	Registraion	
10:00-10:30	Guidance	1. opening speech (Organizer) 2. guidance of workshop's flow (Authors)
10:30-12:00	Lecture session	1. Why do we use the nuclear power generation at all? (Speaker: Agency for Natural Resources and Energy) (25min.) 2. Structure and Safety of geological disposal (Speaker: NUMO) (25min) 3. Promoting process of HLW disposal project and Symbiotic system with the siting areas (Speaker: NUMO) (25min)
12:00-13:00	(Lunch)	
13:00-15:30	Question-and-answer session	0. Information booth (Authors) 1. Nuclear energy booth (Agency for Natural Resources and Energy) 2. Structure & Safety booth (NUMO, expert) 3. Promoting process & Symbiotic system booth (Agency for Natural Resources and Energy, NUMO)
15:30-16:00	Opinion formalization session	
16:00-17:30	Communication session	Facilitator of this session: Organaizer
17:30	Closing	1. Closing speech (Organizer)

Fig. 1 shows the design model in the question-and-answer session. We used four rooms on this workshop. These four rooms are divided, but each room has large window. So we hoped that participants may feel easy to move four rooms freely. However, they divided three groups and moved three rooms in turns basically on contrary of authors' thinking. Some of participants moved freely or came to the information booth to be advised.

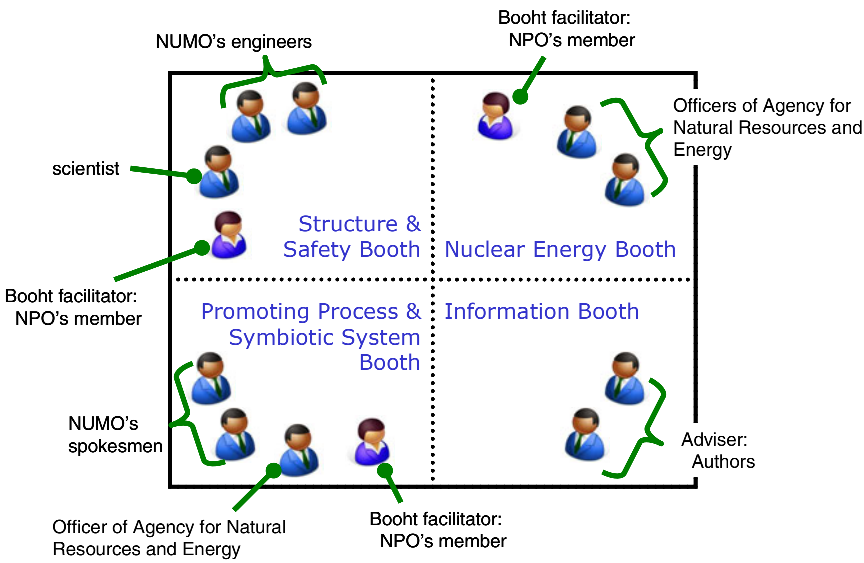


Fig. 1. Design model in Question-and-answer session

4 Participants' Opinions after the Workshop

After the workshop, we conducted interviews to some participants in order to know what effects they have through the workshop. In this interview, we asked participants' opinions concerning HLW disposal.

In the results, we found that the participants provided opinions on concerning three topics, the safety of HLW disposal, the political process to seek voluntary areas, and about the government, local government and promoting body. We introduce about the opinions of each topic as follows.

The Safety of HLW Disposal. Participants would like to know about the HLW disposal in detail, under the condition that they could understand that the disposal is necessary. If it is difficult for the interviewees to understand the safety of the HLW disposal in detail, they may base their conclusions based on trust to the speakers who explain the safety, or the opinion leaders of his or her town. If participants can trust to the speakers or opinion leaders, they may trust their messages. Even if Interviewees can understand (or trust) the safety of the HLW disposal, their vague anxiety is not swept away completely.

Political Process to Seek Voluntary Areas. Most of the participants typically reached opinion that the HLW disposal project cannot be promoted using the open solicitation method to seek siting areas. The participants suggested that, at first, the government and promoting body should select the best place to dispose of HLW and ask for understanding from the residents living there with sincerity and enthusiasm, instead of the open solicitation.

Government, Local Government and Promoting Body. It is necessary for the government, the promoting body or local government to be prepared to achieve HLW disposal clearly. When the participants consider whether the HLW disposal project is applied in their town, they call into account the organizational culture of the government and the promoting body, such as the sense of sectionalism, high pace of change of position, the negative image in the nuclear industry segments, etc.

5 Summary and Future Work

In Japan a disposal of HLW is a pressing issue. Especially, the process of selecting the siting areas for the HLW geological disposal proceeds with difficulty. One of reasons of this difficulty may be that many people don't have enough knowledge about HLW disposal to evaluate the issue completely.

In this research, we introduce one of communication field design in order to lead to satisfied understanding concerning HLW disposal. This workshop consists of four sessions; a lecture session, a question-and-answer session, an opinion formalization session, and a communication session.

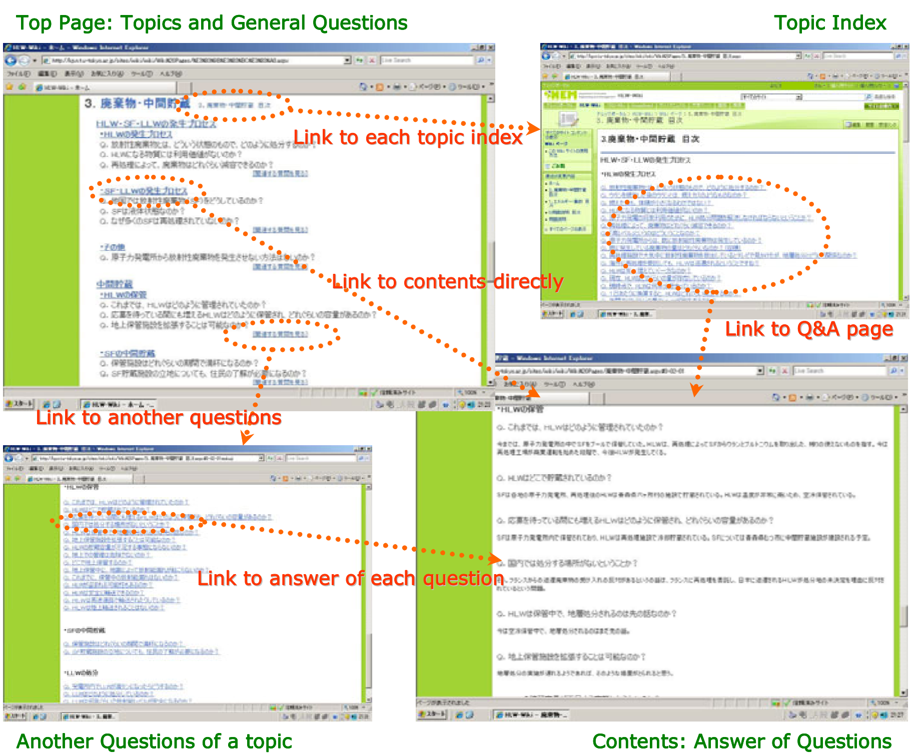


Fig. 2. One shot of online wiki system concerning HLW disposal (now developing)[3]

This workshop was conducted on June 20th, 2010 in Tokyo, Japan. Twenty citizens participated in this workshop. All of participants are aware of HLW, and almost of them had learned about HLW disposal shortly before the workshop. The experts included a scientist, NUMO's engineers, NUMO's spokesmen, officers of Agency for Natural Resources and Energy. The information advisors were handled by the authors. Additionally we commissioned the professional of facilitation for the workshop's facilitator.

After the workshop, we conducted interviews to some participants in order to know what effects they have through the workshop. In the results, we found that the participants provided opinions on concerning three topics, the safety of HLW disposal, the political process to seek voluntary areas, and about the government, local government and promoting body.

Through this workshop, we may know important factors in order to lead many people to understand about the HLW disposal. Now, based on these results, we are developing the data base system about HLW disposal using wiki system. Fig. 2 is the one shot of this developing system. we hope this system to lead to communicate between experts and non-experts on the internet and satisfied understanding concerning HLW disposal.

References

1. Flynn, J., et al.: Trust as a determinant of opposition to a high-level radioactive waste repository: Analysis of a structural model. Risk Analysis 12, 417 (1992)
2. Kimura, H.: Comparative Study Of Cognition Concerning High-Level Radioactive Waste Between Nuclear Experts And Non-Experts. In: 19th International Conference on Nuclear Engineering, Chiba, Japan, May 16-19 (2011) (accepted)
3. Sugiyama, D., et al.: Development of Risk Communication Methodology concerning Safety Case of Radioactive Waste Disposal, CRIEPI Report L980804 (2009) (Japanese)

Control Error Analysis of Computerized Operational Environment in Nuclear Power Plants

Seung Jun Lee, Jaewhan Kim, and Seung-Cheol Jang

Korea Atomic Energy Research Institute,
1045 Daedeok-daero, Yuseong-gu, Daejeon
305-353, Republic of Korea
{sjlee,jhkim4,scjang}@kaeri.re.kr

Abstract. The designs of advanced main control rooms (MCRs) in recently constructed or developing nuclear power plants (NPPs) have been changed by adapting digital and computer technologies. Conventional display systems and controls have been replaced with computerized displays (large display panels and computer screens) and soft controls in advanced MCRs. These different interfaces require operators to perform different tasks for operating and maintaining NPPs. The different operational environment could cause new types of human errors or can vary the possibility the errors. In this work, human errors which could occur during operation executions using soft controls were analyzed. This work suggested human error modes for soft controls and a quantification method for estimating possibilities of the human errors.

Keywords: human error, HRA, main control room, nuclear power plant.

1 Introduction

Human error is one of the main causes of nuclear power plant (NPP) accidents and there have been many researches about human reliability analysis (HRA) in its maintenance and operations [1], [2], [3]. As digital and computer technologies have been adapted in NPPs, there have been many changes in the design of main control rooms (MCRs) [4], [5]. Advanced MCRs have been designed based on much simplified interfaces using large display panels, computerized displays, soft controls, computerized procedure systems and so on. Due to these different interfaces, different human errors should be considered in the HRA for advanced MCRs. In this work, human errors which could occur during operation executions using soft controls were analyzed.

1.1 Advanced Main Control Rooms in Nuclear Power Plants

As digital and computer technologies have been adapted in NPPs, there have been many changes in the design of MCRs [4], [5]. While there are lots of alarm panels,

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 360–367, 2011.

analog indicators, and hand switches in conventional MCRs, advanced MCRs have been designed based on much simplified interfaces using large display panels (LDPs), computerized displays as shown in Fig. 1. In advanced MCRs, the operation environment has considerably changed compared with that of conventional MCRs. For example, operators in advanced MCRs do not have to move to view the plant variables. They can search for the necessary information by navigating computerized displays and control devices using the mouse or tough panels at their positions. These kinds of computerized displays and soft controls may make operations more convenient but they can cause new types of human errors. The computer-based interface could cause the increase of time necessary for the searching of information and the possibility of the viewing wrong indicators. Also, the actions for the operations are performed by soft controls in advanced MCRs. While the predominant means for providing control input is via hard-wired, spatially dedicated devices that have fixed functions in conventional MCRs, the operator may interact with the plant via "soft" controls in advanced MCRs which are designed by computer-based technologies [6].

Fig. 1. Main control room of APR1400 [7]

1.2 Characteristics of Soft Controls

The actions for the plant operation are performed by soft controls in advanced MCRs. These are control devices having connections with control and display systems that are mediated by software rather than direct physical connections. Consequently, their functions may be variable and context dependent rather than statically defined. For example, a particular control action may produce different results based on the mode of the soft control. Also, devices may be located virtually rather than spatially dedicated.

That is, personnel may be able to access a particular soft control from multiple places within a display system[6]. Fig. 2 shows the soft control of APR-1400 MCR[7]. In order to control a device, an operator searches an appropriate screen including the target device, selects the device, and executes an operation on the pop-up window for the control in this MCR.

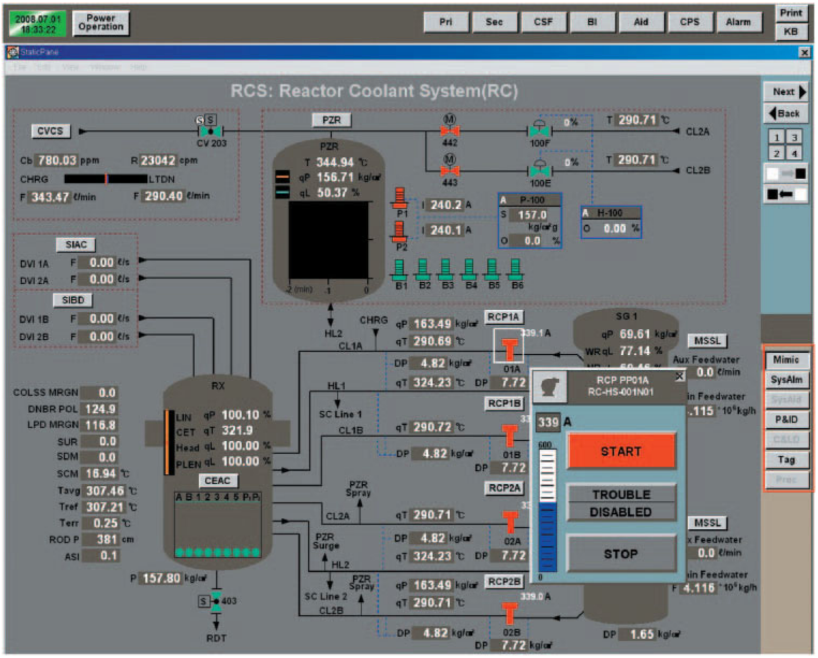


Fig. 2. Computerized display of APR1400 MCR [7]

There are various kinds of input devices for the soft controls: tough screen, light pen, mouse, trackball, joystick, and so on. Because the pace of innovation for computer-based technologies is rapid, this should not be considered an exhaustive listing.

NUREG/CR-6635 explains the general characteristics and definitions of human errors which can occur in soft controls [6].

- Multiple locations for access – Every control in a conventional MCR has a specific location and is connected with a specific plant variable. However, a soft control for a particular variable can have many locations in an advanced MCR. Plant variables are stored in computers and they may be accessed from more than one display device, and from multiple display pages within a display.
- Serial access - Conventional controls provide parallel access. All are visible at the same time. Thus, operators can visually scan controls to observe their status. Advanced MCR components usually contain more displays and controls than can be viewed at one time via the display devices.

- Present and available- Conventional controls are spatially dedicated, and therefore continuously present in the MCR. Soft controls may be designed to be continuously present, or to be retrieved from a display system when needed. In the latter case, they may be considered to be available but not present. In addition, the availability of soft controls may be restricted to specific conditions.
- Physical decoupling of input and display interfaces - Conventional controls typically have closely coupled input and display interfaces. That is, operators perform the input actions and monitor feedback at the same location. However, with soft controls the location of control action may not be closely coupled with the presentation of feedback; the operator may take a control action in one place and read the setting elsewhere.
- Interface management control - In advanced MCRs, operators typically navigate displays and carry out retrieval actions to access them. Thus, actions controlling the user interfaces can be distinguished from actions that control the plant. These two types of actions may be undertaken with the same or different input and display devices.
- Multiple modes - While a conventional control typically performs a single control function, a soft control may perform a range of them each representing a different mode of a soft control device. A specific action, such as pressing a button, can produce different results depending on the different modes as the particular display page currently accessed, the status of the control system, and the status of the plant.
- Software-defined functions - Because operators' actions are interpreted by software, many operations may be initiated via a single action using a soft control. While conventional control systems also have this capability to some degree, software-defined functions can allow systems designers to develop more complexly linked operations.
- Interface flexibility - Computer-based technology can allow the user interface of soft controls to be adapted to changing needs or conditions of use.

2 Identification of Human Errors in Soft Controls

In this work, human errors which could occur during the operations using soft controls are analyzed. For simplicity, diagnosis processes and other additional actions for operations such as searching reference information are not considered.

In spite of the different features of advanced MCRs, several human error modes are the same as those of conventional MCRs. However, some human error modes should be considered as more important and additional error modes may be needed because of special features of soft controls in advanced MCRs.

Generally, sequential tasks are performed in soft controls in order to perform an operation. The plant information of advanced MCRs is provided to operators by computer screens as hierarchical forms due to the spatial limits. While device controllers are widely spread and located in different positions in conventional MCRs, operators need to navigate the screen and to select the target device in advanced MCRs. The operation actions of operators may be related to primary tasks, such as providing control inputs to plant systems, or secondary tasks, such as manipulating the user interface to access information or controls or to change control modes.

Interface management tasks are referred to as secondary tasks because they are concerned with controlling the interface rather than the plant. Operators should perform secondary tasks to find appropriate screens or devices by screen navigations and selections before they perform the primary task to control a device.

When operators are required to control a device using soft controls, they have to do four sequential actions in advanced MCRs. First, operators decide what operation should be performed for the current situation. Then, they have to navigate the screens in order to find the appropriate screen having the target device. Then, the device should be selected by a mouse click or other selection methods on the screen. And finally, they control the device in the control panels. This process can be different according to the interface of soft controls.

- Operation selection: according to the operating procedures, an operator selects an operation which is appropriate for the current situation.
- Screen selection: an operator navigates the screens in order to find the target control device. Only one navigation or two or more navigations could be required.
- Control device selection: after selecting the appropriate screen including the target control device, an operator selects the device by pointing input devices.
- Operation execution: an operator performs the required operation on the device.

The operations using soft controls are performed in such a process, and possible human errors during the process could be classified into six types as shown in Fig. 3, as follows:

- Operation omission: an operator has a potential to omit a necessary operation when he/she selects an operation to be performed in operating procedures. Also, after selecting the right operation and selecting the right control device, the operation could not be performed due to an inappropriate manipulation of soft controls.
- Wrong object: when an operator selects the target device, he/she could select a wrong device. In this case, if the operator recognizes the wrong selection, then the operation could be recovered. However, if he/she does not recognize that the selection is wrong and continues executing the operation on the wrong object, then a wrong operation will be performed.
- Wrong operation: even though an operator performed appropriate navigations of screens and right selection of the target device, an operator can perform a wrong operation such as pressing ‘OPEN’ button instead of ‘CLOSE’ button.
- Mode confusion: if a control window includes multi-mode, an operator could perform an operation on the wrong mode. An operator can make a mistake such as increasing the level of pressurizer instead of its pressure in case that the level and pressure of pressurizer are controlled in the same control panel with mode switch.
- Inadequate operation: when an operator executes a right operation after appropriate navigations of screens and right selection of the target device, the operation could be executed insufficiently, too early or too long/short. All operations which are performed incompletely and inappropriately belong to this error mode.
- Delayed operation: due to the wrong selections of screens or devices and recovery of them, an operation could not be performed on the right time. Additional time for re-selection of screens or devices could be one of the reasons of such delayed operation.

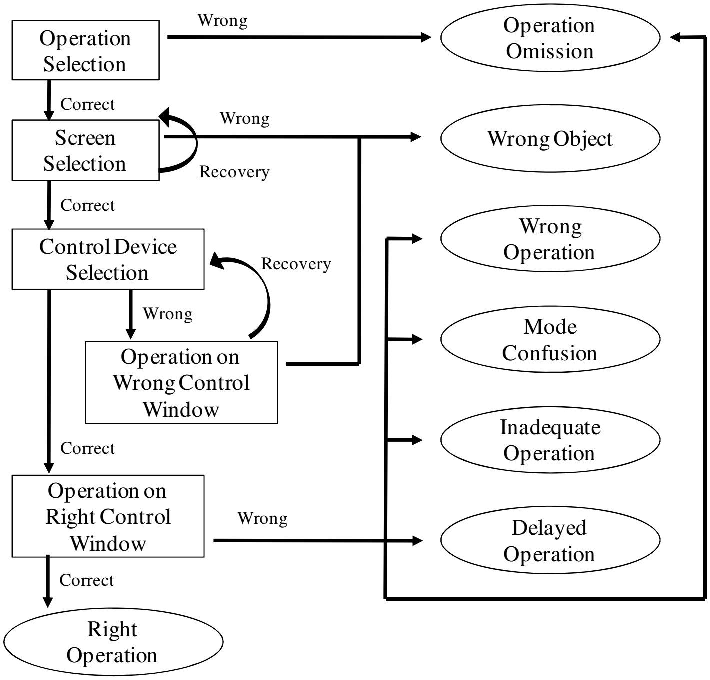


Fig. 3. Process and possible human errors of soft controls

3 Quantification of Human Errors in Soft Controls

In this work, the possibility of human error was quantified based on K-HRA method. Human error probability of K-HRA is calculated by multiplication of complexity of a task and recovery failure probability of an error. Also, the complexity of a task is determined by stress level of an operator and type of the task. Some factors of soft controls may have almost similar values compared to those of conventional ones (e.g., stress level of an operator) because those factors are not much affected by different interface designs. However, some factors (e.g., recovery failure probability) should have specific values for soft controls. The control processes of soft controls and conventional ones have different sub-tasks (e.g., screen navigations are not necessary in conventional controls). Moreover, all operators in control rooms can share operational information through computerized displays and useful functions using computerized interfaces could be provided for preventing human errors. This environment could cause increase of recovery probability or require more workload of operators.

In this work, a quantification method for human errors in soft controls was proposed, which is called d-HRA. The equation for estimating execution errors of d-HRA is

Pr(Control error) = Σ [Pr(Error of unit task)) * Pr(Recovery failure)]
* f(Interface management complexity)
Pr(Error of unit task) = f (Task type, Stress level)
Task type = f (Unit task complexity, Quality of procedure, Expertise for the task)
Stress level = f (Time urgency, Situation severity, Task risk, Education/Training)
Pr(Recovery failure) = f (Time urgency, Quality of MMI, Check/Confirmation)

The same parameters and values are used in the d-HRA method for some factors (e.g. stress level and task type). In some parameters, the levels and values of them were changed due to the features of soft controls.

- Quality of MMI (Location of feedback): Due to the flexible design of advanced MCR interfaces, indicators for feedbacks can be located in not fixed location. For the quality of MMI for location of feedback, two aspects are considered: 1) existence of direct feedback indicator, 2) location of the indicators of the feedback. For example, if there is an indicator which directly represents the result of an action and it is located close to the related control in the same control panel, then the quality of MMI is high. On the other hand, if an operator has to consider two or more indicators in order to see the results of an action and interface managements are necessary, then the quality of MMI is low.
- Check/confirmation (Function for check/confirmation by other operators): Several computerized procedure systems provide a function for check/confirmation of an action by other operators. If there is no function and procedure for check/ confirmation of other operator, then the level of this factor is low. If there is a function for check/confirmation and all actions are performed after check/ confirmation by other operators, then the level of check/confirmation is high.

A new factor, interface management complexity (IM_C), was added in order to consider the special features of soft controls such as navigating screens and handling several kinds of input devices. The consistency among sub tasks in a task and required interface managements are reflected in this factor.

For the quantification of the human errors of soft controls, the effect of each factor of soft controls was determined by experts. They were required to determine the values of factors compared to those of conventional controls. Table 1 shows an application result. 'K-HRA' means the control error probabilities using conventional

Table 1. An application result

Accident number	**K-HRA**	**d-HRA**		**Remarks**
		Check/Confirmation: High	Check/Confirmation: Mid	
1	1.0E-1	1.2E-1	1.2E-1	IM_C: Mid
2	1.5E-1	5.0E-2	1.0E-1	IM_C: Low
3	4.0E-2	2.0E-2	4.0E-2	Local operations
4	5.0E-4	4.50E-4	9.0E-4	IM_C: Mid
5	2.0E-3	1.5E-3	3.0E-3	IM_C: Mid
6	1.0E-3	5.0E-4	1.0E-3	IM_C: Low
7	1.1E-3	8.4E-4	1.7E-3	IM_C: High
8	5.0E-3	3.0E-3	6.0E-3	IM_C: Mid

controls. 'd-HRA' indicates the control error probabilities using soft controls according to the 'check/confirmation level'. IM_C level is shown in remarks field. In accident 3, IM_C is not available because all control actions are performed in local. The results showed that soft controls had both positive and negative effects on human errors according to tasks, situations and quality of the system.

4 Concluding Remarks

In this work, human errors which could occur during operation executions using soft control were analyzed. Soft controls of advanced MCRs have totally different features from conventional controls, so that they may have different human error modes. When operators are required to control a device using soft controls, they have to do four sequential actions in advanced MCRs: operation selection, screen selection, control device selection, and operation execution. This work classified the human errors in soft controls as six types; operation omission, wrong object, wrong operation, mode confusion, inadequate operation, and delayed operation. A quantification method for soft control errors was proposed. An application for the quantification method was performed based on the experts' opinions and the results showed that soft controls had both positive and negative effects on human errors according to tasks, situations and quality of the system. Since one of the features of computerized MCRs is interface flexibility, the analysis of human errors can be used for designing more reliable interfaces and for reducing human errors by appropriate training.

Acknowledgments. This work has been carried out under the Nuclear R&D Program supported by MEST (Ministry of Education, Science and Technology). (Grant code: 2010-0001031).

References

1. Swain, S.D., Guttmann, H.E.: Handbook of Human-Reliability Analysis with Emphasis on Nuclear Power Plant Application; Final report, NRC FIN A1188, NUREG/CR-1278 (1983)
2. O'Hara, J.M., Brown, W., Nasta, K.: Development of the human-system interface design review guideline (NUREG-0700) Revision I (BNL technical report L-13 17-2-12/96). Brookhaven National Laboratory,Upton, NY (1996)
3. O'Hara, J.M., Hall, R.E.: Advanced control rooms and crew performance issues: Implications for human reliability. IEEE Transactions on Nuclear Science 39, 919–923 (1992)
4. Lee, S.J., Seong, P.H.: Development of an integrated decision support system to aid cognitive process of operators. Nuclear Engineering and Technology 39, 703–717 (2007)
5. Lee, S.J., Kim, J., Jang, S.C., Shin, Y.C.: Modeling of a dependence between human operators in advanced main control rooms. Journal of Nuclear Science and Technology 46, 424–435 (2009)
6. Brookhaven National Laboratory. Soft control: Technical basis and human factors review guidance (NUREG/CR-6635). U.S. Nuclear Regulatory Commission, Washington, DC (2000)
7. Lee, M.S., Hong, J.H., Suh, J.K., Lee, S.H., Hwang, D.H.: Development of human foactors validation system for the advanced control room of APR1400. Journal of Nuclear Science and Technology 46, 90–101 (2009)

uMeeting, an Efficient Co-located Meeting System on the Large-Scale Tabletop

Jie Liu and Yuanchun Shi

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

Abstract. In this paper, we present the uMeeting system, a co-located meeting system on the large-scale tabletop. People are used to sitting around a table to hold a meeting. It is natural and intuitive. The table has a central role to support team activities. Horizontal surfaces rather than vertical ones have inherent features to support the co-located meeting. Existing tabletops aren't large enough to support more than 4 people's meeting and the display area for each person is not large enough. Thus we develop uTable, a large-scale multi-touch tabletop. Based on uTableSDK we developed earlier, we develop uMeeting system that supports co-located meeting on the large tabletop uTable.

Keywords: co-located, collaboration, meeting, large-scale tabletop.

1 Introduction

Horizontal and vertical surfaces are two main kinds of interactive surfaces. In the past twenty years, much research has been done in vertical wall-like surfaces for co-located collaboration, like Liveboard[1] and CityWall[2]. If we simply apply these vertical surfaces into the meeting scenario, many problems may emerge. Firstly, the attendees have to stand in front of the surface at the meeting; so vertical surfaces are not suitable for long-time discussion. Secondly, attendees have to turn around to communicate with others by side for eye contact. It lacks face-to-face interaction, which largely reduces the efficiency of communication in the meeting. Last but not the least, the upper part of large-scale vertical surfaces usually is out of reach. So the actual available area for interaction is much smaller than the display area, which decreases the usability of the vertical surface greatly.

On the other hand, horizontal surfaces have inherent features to support the co-located meeting. People are used to sitting around a table to hold a meeting. It is natural and intuitive. The table has a central role to support team activities. Indeed, 55% of all documents and intermediary objects used during a working session are displayed on the meeting table[10]. The table remains the principal collaborative space. Horizontal surfaces use the table as a metaphor and are also known as tabletops. Tabletops are natural interaction substrates for people to meet around and collaborate on. It can easily solve the problems that vertical surfaces cannot solve as mentioned above.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 368–374, 2011.

Existing tabletops such as Microsoft Surface[3] and DiamondTouch[4] aren't large enough to support more than 4 people's meeting and the display area for each person is not large enough. Thus we develop uTable, a large-scale multi-touch tabletop. uTable has an interactive surface of 3.8m x 2m, 3840 x 2048 pixels. It is composed of ten unit tabletops. uTable supports simultaneous multi-touch and pen interaction, so It allows multimodal interaction and can support about 12 people's meeting.

Application systems on existing tabletops focus on general collaboration and don't feature the meeting scenario. Besides, these systems were designed for small group collaboration, and may not fit for large group meeting. We need to design a new meeting system based on uTable platform. Based on uTableSDK[7] we developed earlier, we develop uMeeting system that supports co-located meeting on the large tabletop uTable.

In this paper, we introduced related work about tabletops and co-located collaboration in section 2. We also give a brief introduction about uTable platform and uTableSDK we developed. In section 3, we explain the uMeeting system in detail and list several advantages this system holds. At last, we draw conclusions and show some future work.

2 Related Work

The study on interactive tabletops has a long history. It can be traced back to 1990s when Pierre Wellner introduced DigitalDesk[8]. It projects electronic images down onto the desk and paper documents. People can interact with it using pens or bare fingers. Fitzmaurice et al.[9] used physical artifacts called bricks to manipulate virtual objects on tabletops. This essentially new input device can be tightly coupled or attached to virtual objects for manipulation or for expressing action. In recent years, many tabletops have come into being. Some well-known tabletops are Surface, SMART Table and DiamondTouch, just to name a few. All the tabletops mentioned above share the same drawbacks that they are not so big enough as to afford many people simultaneously. Very large display and interactive area is in great need for large group collaboration.

Large display has many advantages. Ball et al.[11] reported an observation analysis of the use of a large tiled display and concluded that large displays can improve user performance for task switching or viewing large documents, increase ability to spatially position applications and shortcuts for quick access and recall, and can help work collaboratively. But large displays often have bezels, that are the covering around the edges of the display screens which distort the image and affect users' experience largely. When we build uTable platform, we take this into account and try to avoid the bezel effect. Thanks to the technology of Lanbo Corp.[5], the width of joint of unit tabletops is less than 2mm, so the display image and interactive input are seamless and interior bezels effect[6] can be ignored. Hence the quality of display on uTable is great as shown in Fig. 1 and visual and perceptual immersion on uTable is excellent.

Existing interactive surfaces with large display are almost vertical surfaces like screen walls. CityWall[2] is a big interactive wall installed in public circumstance.

CityWall's large physical size appeared to support making interactions visible to others both gestural and as effects on the display when this was wanted. The author also indicated that large display size and visibility also supports immediate availability of content to interact with. But vertical interactive surface isn't the first choice for collaboration for the reasons listed above. When people want to have a discussion, they tend to have a free talk, seated with eye contact. Horizontal interactive surface is preferred when acting as a meeting table. Meanwhile, some compromise and hybrid interactive surfaces are also introduced. Tilted Tabletops [12] combined vertical surface and horizontal surface. The vertical surface is considered to be public whereas horizontal display spaces let users have more privacy. Tilted tabletops provide the opportunity to collaborate side-by-side as well as face-to-face.

Much research on co-located collaboration with tabletops has been made. Scott et al.[13] analyzed digital tabletop systems and suggested several design guidelines for effective co-located collaboration around a tabletop display including supporting interpersonal interaction, supporting transitions between personal and group work, supporting simultaneous user actions and supporting fluid transitions between activities. When we design the uMeeting system, we also take into these guidelines into consideration. Other issues such as the orientation of each workspace and the ownership of workspaces should also be addressed. Scott et al.[14] conducted two observational studies of traditional tabletop collaboration and revealed that collaborators use three types of tabletop territories to help coordinate their interactions within the shared tabletop workspaces: personal, group and storage territories. Comprehensive reports about user experiences about co-located group collaboration also come out.

Fig. 1. uTable platform

We specify the meeting scenario, one of the common collaboration scenarios, to implement on uTable platform. uMeeting system can substitute for laptops, notebooks and printed materials. We classify the territories into three scopes, namely public, private and group scope. Presentation is a typical public scope application. Taking notes and anonymous vote are in private scope. Activities such as discussion, sharing digital material and making up a schedule are in group scope.

uTable is a very large tabletop we build. As shown in Fig. 1, uTable is tiled seamlessly. The surface is rear projected and multi-touched. To summarize, uTable has three main features: 1) very large interactive surface, support many users seated; 2)Tiled, seamless both in display and interactive input; 3) support simultaneous multi-touch and pen interaction based on diffused illumination.

To enable developers rapidly develop customized application on uTable, we also developed uTableSDK[7], a software development kit for uTable. uTableSDK provides many practical features, such as multi-users, multi-touch input handling, UI management policies on large tabletops, personal space management and communication approaches between each other. We use uTableSDK to develop uMeeting system with a lot of time and effort saved.

3 System Design and Implementation

To design an efficient meeting system on uTable, we first study what people do when attending a meeting and categorize most frequent activities by the scope of message transmission. There are three scopes, public, private and group scope. When attending a meeting, attendees can be seated at will and has their own workspace. The workspace can be customized. After a default workspace is presented, attendees can resize, rotate and choose preferred style of it. If an attendee moves around, the workspace should follow him where he is. The system should also support task migration, which is one same task can be transferred from one workspace to another workspace just like a pipeline. For presentation, the system must have the function of broadcasting. When someone is giving a presentation, all the workspaces should show the same content. For group discussion, several attendees share the same screen and can also have their own area.

To put all these features into the uMeeting system, we need a well-structured and scalable framework. The framework is shown in Fig. 2 below. The architecture of uMeeting system is much like MVC (model, view and controller) style. For every attendee, there is an independent view for him. Attendees can get all information from their own view and interact with the view. The view gets the input instruction and parses it, then sends it to the controller. The controller takes charge of the communication between the user and the system. The controller keeps track of users' information and the tasks that they are running and issues requirement to the model. The model holds all the functions the system has and manages the all the tasks. When a requirement comes, the model processes it and returns a notification.

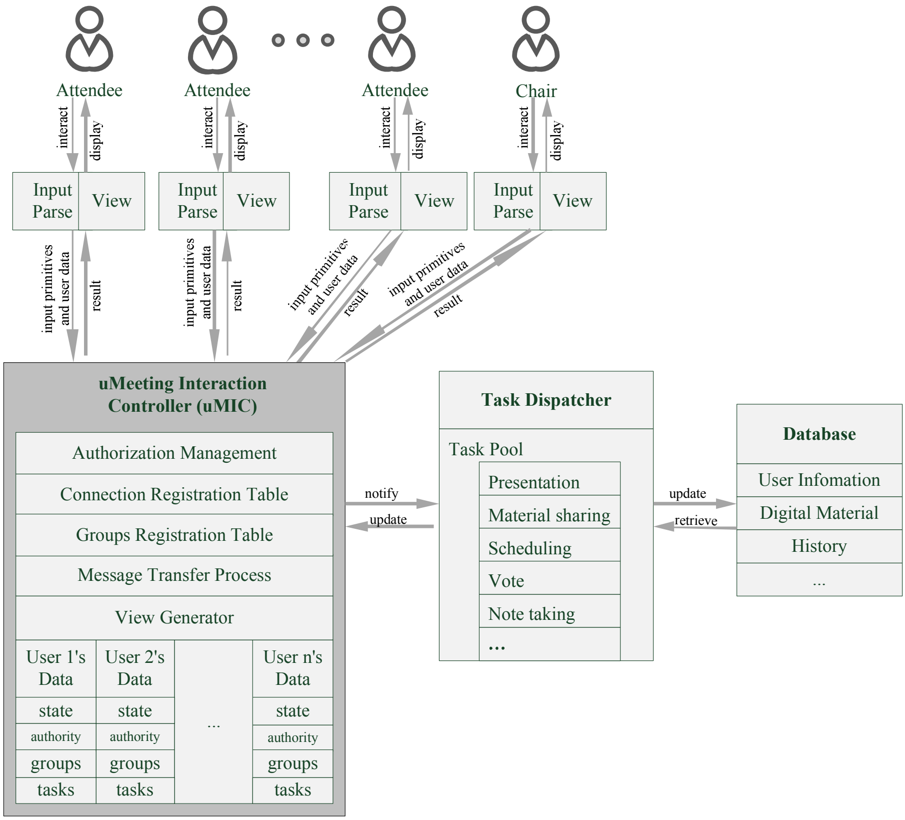


Fig. 2. uMeeting architecture

We emulate the process of the meeting and the roles of each attendee plays. The chair in the meeting has the administration privilege, and can conduct the progress of the meeting. The system creates a stack for every attendee who logs into the system after authentication. In each attendee's stack, there are state and authority information, group information that which this attendee belongs to, and tasks that this attendee is running. Based on these stacks, uMeeting Interaction Controller (uMIC) keeps track of all activities running on the system. Attendees interact with the system by multi-touch, which is a direct and intuitive interaction modal. Input Parse module processes these interaction signals and send out input primitives which can be interpreted as instructions by uMIC. uMIC updates some registration tables and notifies Task Dispatcher to implement these instructions. Then Task Dispatcher selects a task from the task pool and allocates resources to the task to processes the instructions. After completion, results return to uMIC. Then uMIC updates registration tables and states, mixes the result with some visual effect and transfers these data to the View module. At last, attendees see the result.

There are many advantages in the design of uMeeting system. First, attendees can be seated flexibly without restriction. Attendee simply creates a personal stack where he or she is seated. By using task pool and user's data stack in uMIC, one kind of task needs only one process to handle it. At the same time, each attendee who is running this task can get a different view of the result. Moreover, task migration in collaboration is simple since there are no consistency problems. We use Groups Registration Tables to separate public, private and group scope. Private scope can be treated as a group of the attendee himself. Message Transfer Process module transfer message among each group and View Generator generates views for each attendee. Besides, if a group of attendees want to share a big view, View Generator can join the views and generate one view for this group.

Fig. 3. One scenario when using uMeeting system

4 Conclusions and Future Work

In this paper, we have presented the uMeeting system, a co-located meeting system on the large-scale tabletop. First, we compare the horizontal interactive surfaces and the vertical ones and draw the conclusion that horizontal surfaces are preferred for the meeting scenario. Then we list some activities that attendees do when attending meetings. We classify the workspace into three scope, that is public, private and group scope. For each scope, we list some specific activities. We implemented the uMeeting system based on uTableSDK and gave a detailed explanation about the architecture and advantages of the uMeeting system.

Future work includes conducting a comprehensive user study to find to how much degree this meeting system helps to facilitate the progress of a specific meeting. Which category of activities in the meeting helps attendees a lot and which helps little. With the changing of devices used in meeting from notebooks and laptop computers to large-scale tabletop, the impact on the attendees' behavior should also be studied.

Acknowledgements. This work is supported by National High-Tech Research and Development Plan of China under Grant No.2009AA01Z336.

References

1. Elrod, S., Bruce, R., et al.: Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations and Remote Collaboration. In: Proc. of CHI 1992, pp. 599–607 (1992)
2. Peltonen, P., Kurvinen, K., Salovaara, A., et al.: It's Mine, Don't Touch!: Interactions at a Large Multi-Touch Display in a City Centre. In: Proc. of CHI 2008, pp. 1285–1294 (2008)
3. Microsoft Surface, `http://www.microsoft.com/surface/en/us/default.aspx`
4. Dietz, P., Leigh, D.: DiamondTouch: A Multi-User Touch Technology. In: Proc. of UIST 2001, pp. 219–226 (2001)
5. Lanbo Corp, `http://www.bjled.com/`
6. Bi, X., Bae, S.H., Balakrishnan, R.: Effects of Interior Bezels of Tiled-Monitor Large Displays on Visual Search, Tunnel Steering, and Target Selection. In: Proc. of CHI 2010, pp. 65–74 (2010)
7. uTableSDK, `http://utablesdk.codeplex.com`
8. Wellner, P.: Interacting with paper on the DigitalDesk. Commun. ACM 36(7), 87–96 (1993)
9. Fitzmaurice, G.W., Ishii, H., Buxton, W.: Bricks: Laying the Foundations for Graspable User Interfaces. In: Proc. of CHI 1995, pp. 442–449 (1995)
10. Laborie, F., Jacquemond, D.: Cooperation in Highly Distributed Design Processes: Observation of Design Sessions Dynamics. In: Proc. of CDVE 2005, pp. 42–51 (2005)
11. Ball, R., North, C.: An analysis of user behavior on high-resolution tiled displays. In: IFIP International Conference on Human-Computer Interaction, pp. 350–364 (2005)
12. Tomfelde, C.M., Wessels, A., Schremmer, C.: Tilted Tabletops:In Between Horizontal and Vertical Workspaces. In: Proc. of TABLETOP 2008, pp. 49–56 (2008)
13. Scott, S.D., Grant, K.D., Mandryk, R.L.: System Guidelines for Co-located, Collaborative Work on a Tabletop Display. In: Proc. of ECCSCW 2003, pp. 159–178 (2003)
14. Scott, S.D., Sheelagh, M., Carpendale, T., Inkpen, K.M.: Territoriality in collaborative tabletop workspaces. In: Proc. of CSCW 2004, pp. 294–303 (2004)

Enhanced User Experience in Managing Personal Finance

Cindy Lu

iNautix USA, User Experience Designs, United States
clu@inautix.com

Abstract. This paper reviews several online personal finance management tools to understand if rich interaction designs and technologies have added values to help individuals better manage their personal finance. Goals, tasks and user experience expectations in managing personal finance are discussed. The interaction designs and impact on user experience in selected online financial management tools are summarized. Some negative impact is also noted. Finally, it discusses the challenges and suggests opportunities to user experience professionals to address in the future.

Keywords: RIA, Rich Interactions, Interaction Designs, User Experience, Personal Finance, Web Applications, Online Tools.

1 Introduction

One of the challenges in life for individuals and families is to manage personal finance. It is an overwhelming task not only because people have to learn the knowledge and skills in many specialized areas such as banking, credit cards, investments and financial planning but also they have to learn and use various tools that have different ways in presenting information and interactions.

Over the last 10 years, many online tools have become available for managing personal finance. People can use these tools for almost everything from financial planning, budgeting, banking, and paying off debt to managing retirement and investments. They can also use consolidation tools to monitor overall assets, investments, spending, and analyzing patterns and trends.

Businesses have put great effort to enhance the tools to make it easier for average people to use. For example, the look and feel of online applications is consistent even though the data is pulled from different backend systems, users can single sign-on to related applications, wizards are used to guide users in complex tasks such as financial planning or making investment decisions, and contextual help is presented to users to understand terminologies. Rich interaction designs and related technologies have been applied to many financial management tools to improve the user experience. Old personal finance tools and web application have been enhanced with richer information and interactions and new tools have come up.

Have rich interaction designs and technologies enhanced user's experience in using personal finance management tools? If yes, what tasks have been better supported?

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 375–383, 2011.

This paper reviews the user experience designs in selected personal finance management tools to understand if rich interaction designs and technologies have added values to better supported users' goals and tasks, what experience levels have been enhanced and opportunities for further improvement.

2 Managing Personal Finance

2.1 Personal Finance

Personal finance is the application of the principles of finance to the monetary decisions of an individual or family unit [1]. Individuals practice personal finance management aim to live the life they envisioned with limited financial recourses, taking into consideration of various financial risks and future life events.

To manage personal finance, one has to learn the basic principles and terminologies such as assets, liabilities, credits, risks, banking, investments, loans and taxes. Individuals also need to take actions such as creating financial plans and budgets, monitoring expenses, managing investments, paying credit cards and loans,

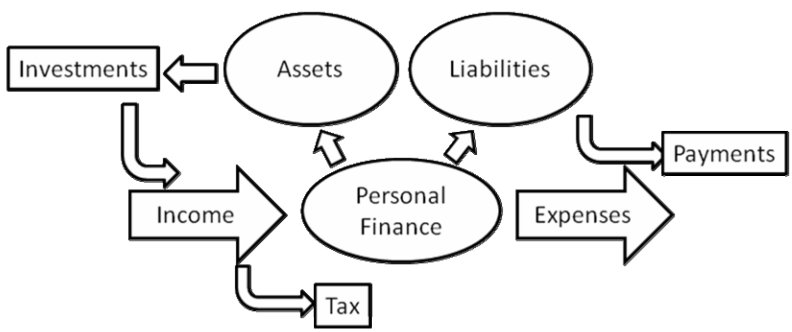


Fig. 1. Personal finance is about managing the personal or family money based on the principles of finance

Table 1. Sample Goals and Tasks in Personal Finance Management

Category	Goal	Task
Plan	To set up a financial roadmap	Create a financial plan
	To balance income/spending/saving	Budget money
Operation	To save for retirement	Manage retirement accounts
	To grow assets	Manage investments
	To understand spending habits	Record expenses
	To pay off debt	Manage loans and credit cards
	To fulfill tax obligations	File income tax
Monitor	To understand and reduce expenses	Analyze spending pattern
	To detect fraudulent charges	Monitor credit card charges
	To ensure financial goals are met	Monitor asset growth and spending

and filing income tax. In general, managing personal finance can be a daunting task especially for a busy family. For individuals who manage personal finance by themselves, the tasks can be broken down to three types, plan, operation and monitor. Table 1 lists sample goals and tasks.

2.2 Tools in Personal Finance Management

There are two types of tools that people can use to manage personal finance: 1) tools that help individuals to manage a specific area such as banking, credit cards, online payment or investments and 2) tools that aggregate data from various sources and provide high level pictures about personal wealth. Table 2 lists sample personal finance tools. The paper reviewed tools in banking, credit cards, investments (Citi Bank, American Express, Ever Bank, ING, Fidelity and TD Ameritrade) and consolidation tools (Mint.com).

Table 2. Sample Personal Finance Tools

Category	Name	Description
Budgeting	YNAB.com	Budget and track expenses
Credit Cards	Citi Bank	Manage credit cards
	American Express	Manage credit cards
Banking	Ever Bank	Manage credit cards
	ING	Manage credit cards
Investment	Fidelity	Manage investments
	TD Ameritrade	Manage investments
	T Rowe Price	Manage investments
Tax	Turbo Tax	File tax
	TaxCut	File tax
Consolidation	Mint.com	Provide overall wealth picture
	Quicken	Provide overall wealth picture

3 User Experience

3.1 User Experience Expectations

Users acquire tools to assist them in achieving specific goals. User experience is about the journey of using these tools. Depending on the usage context and knowledge and skill levels, users have different expectations for using the tools. For example, for playing video games, users may expect to have a fun experience but for managing personal finance, users would expect to feel secure and confident about data and transactions. In general, users want to get the job done and they want to feel good about completing it.

There are several user experience models that describe the elements of user experience and the priority in product design and development. Garrett and Berry both defined user experience elements from product design view point [2] [3].

Morville defined six facets of the user experience honeycomb [4]. Cerejo defined user experience needs based on Abraham Maslow's pyramid of human needs [5]. Several articles discuss Kano model and its impact on user experience [6][7].

Managing personal finance is not a fun task. It not only requires the user to have a good understanding of basic principles but also be disciplined to perform routine and detailed tasks. Users want to get the job done but want to have the assurance about doing it right. The experience expectations can be modeled below:

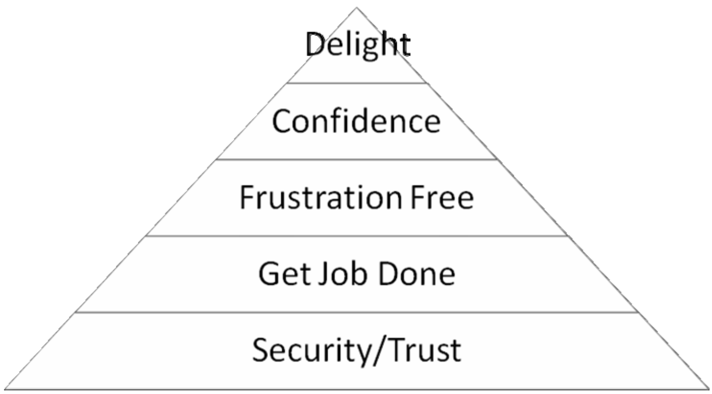


Fig. 2. User Experience Model - Managing Personal Finance

(1) Security/Trust. Users must feel the system is secure and trustworthy. This is especially important in managing personal finance. They want to be assured that the data and transactions are safe and accurate.
(2) Get Job Done. Users hire the system to complete a job. They have expectations on basic functions that can get the job done.
(3) Frustration Free. Users have basic expectations on the fluidity of performing tasks. When these basic needs are not met, they feel frustrated.
    a. Fast. No matter whether it is an experienced or novice user, consumer or professional, people expect to use the tool to complete tasks in a shortest time. Slow system response and multiple steps will cause frustration.
    b. Easy. Users want to figure out things instead of reading instructions or relying on memory. They want to interact with the system with ease and fluidity.
(4) Confidence: Users want to feel confident about the information presented to them and about the actions they have taken. If there is an error, they want to know what it is and how to fix it quickly.
(5) Delight: Even though users do not expect to have a pleasure in performing the tasks in managing personal finance but pleasant visual, interaction designs or additional value-added features will make the user feel good when performing the tasks. Value added information that can help use better understand the data and trends will make people feel delightful.

3.2 A Persona

Kevin Smith, 40 years old, is a relationship manager in a large pharmaceutical company. He manages ongoing contract relationships and service delivery to clients for multiple accounts. Kevin has a family of four, his wife Jane and teenage daughters, Amanda and Peter. Kevin and his wife understand that managing personal finance will enable them to plan for the future, achieve their financial goals and live the life they envisioned.

Kevin and Jane's financial goals include: 1) To pay off the mortgage in 10 years, 2) To send kids to the best colleges they choose in 5 and 6 years respectively 3) To remodel the kitchen in 2 years, and 4) To live comfortably when retiring between ages 60 – 65. Kevin performs following regular tasks (see Table 3) to make sure their financial is on target.

Table 3. Kevin and Jane's Tasks in Managing Personal Financial

Category	Task	Frequency
Plan	Create budgets	1-2/year
	Create goals for savings	1-2/year
Operation	Banking	Weekly
	Research saving tips	Weekly
	Pay credit cards	Monthly
	Pay mortgages	Monthly
	Making investments	Quarterly
	Managing retirement accounts	Quarterly
	File tax return	Yearly
Monitor	Monitor spending	Weekly
	Monitor credit card charges	Weekly
	Monitor investment performance	Quarterly
	Monitor education fund performance	Quarterly
	Monitor net worth and asset growth	Quarterly

Kevin's expectations to the tools managing the personal finance:

1) Secure
2) Data is accurate and up to date
3) Provide whole picture – how much I have and investment performance
4) Save time – easy to find details; enable me to take quick actions; respond to me quickly
5) Information is easy to understand

4 Rich Interaction Designs in Personal Finance Tools

Two methods are used in the tool evaluation, walking through the tasks performed by persona and user experience heuristics. The persona is created based on past studies in online investor projects and analysis of articles and online forums in personal finance.

Goals and sample tasks performed by Kevin and expectations and limitations were used in the evaluation. User experience heuristics includes simplicity, information alignment (is the information presented serves the goals of the user?), conversation/ engagement and appearance and user experience models (Figure 2).

4.1 Rich Interaction Designs and Enhanced User Experience

Frequent tasks performed by the persona, Kevin, were examined in the selected tools. Rich interaction designs and technologies that support these tasks are identified and summarized below. Among the tools surveyed, some tools have applied rich interaction designs for the entire application. Other tools have embedded rich interactions and technologies into some sections. For example, the two online brokerage sites studied embedded interactive graphics into the portfolio and balance viewing, research and screening sections and retirement planning.

4.2 Improved Design and Interactive Patterns

Several information designs and rich interaction design patterns are observed in the surveyed tools. The table below summarized these observations.

4.3 Opportunities for Improvement

The most frequent issue observed in the surveyed tools is the time waiting for page or component to be loaded. When a partial section is being loaded and the user can perform other tasks at the meantime, it OK for the user to view the data loading image. However, when the entire page is being loaded and the user cannot do anything else, it is a frustrating experience.

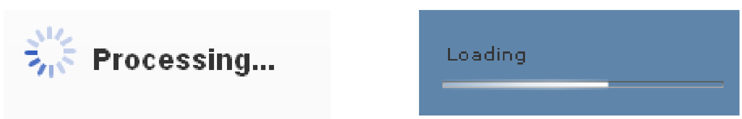


Fig. 3. Sample data loading images

Another problem observed is that when additional information related to the current task is presented to the user, it is not clear where the new information comes from. For example, one application displays US average spending related to the spending category that the user is viewing. It is not clear how the US average is calculated and whether the data is credible. It will be helpful to the user if the source of information is presented to the user also.

One interesting issue is when using interactive charts to enable users to view detailed information or drill down data, how much interactivity it should provide. For example, on one brokerage site where an interactive chart is used to show the user account balance history, interactive control for change timeline is outside of the chart.

Table 4. Frequent Tasks, Rich Interactions and User Experience Supported

No	Task	Rich Interaction	User Experience
1	View Asset Portfolio	- Retrieve information asynchronously and display last update info for each account or asset when the update is done. Kevin can perform other tasks meantime.	- Feeling informed about data update - Efficiency
2	View transactions	- Auto suggest when searching - Instant refresh when searching, filtering and sorting transaction data. No page refresh. - Display spending history for comparison (Average Spending, US) - Filter transaction by customer tags	- Efficiency - Efficiency - Added value - Added value
3	Manage Transactions	- Inline editing - List dynamically built up when entering split transactions - Overlay to display transaction details	- Efficiency - Efficiency - Efficiency
4	View account balance	- Overlay to display account details	- Efficiency
5	View investment performance	- Instant refresh of data and graphics - Interactive chart - Contextual help in video	- Efficiency - Added value - Added value
6	Make credit card payment	- Overlay for displaying charge details - Make multiple payments on one page	- Efficiency
7	Manage Budget	- Create budget on an overlay with background information visible - Graphic view of budgets - Budget comparison - View budget for past 12 months	- Efficiency - Easy to view info - Added value - Function
8	Manage Financial Goals	- Graphic display of goals - New design metaphor for goal creation	- Easy to view info - Easy to enter info - Fun
9	Analyze trends	- Interactive chart that enables date drill down - One click to switch to different category or timeline	- Easy to view info - Efficiency - Fun - Efficiency
10	Research on saving tips	- Instant refresh when filter is changed - Slider bar and text fields for data entry	- Efficiency
11	Set up accounts and preferences	- Overlay with background information visible - Enable multi-tasking, i.e. enable the user to set up other accounts while the data is being loaded	- Efficiency
12	File Tax Returns	- Graphic display of roadmap - Filed highlighter for invitation	- Easy to view info - Easy to view info
13	Help	- Real time information from online community related to the current topic - Quick preview for online posting	- Added value - Easy to view info

Table 5. Most Used Design Techniques in Surveyed Tools

Category	Designs	User Experience Supported
Navigation	- All menu items are visible on the top level navigation. The rest navigations are displayed with content.	- Easy to navigate the main navigation
Information Display	- Minimum text - Minimum number of columns in tables - Plain language. Contextual help for terminologies. - Large font	- Easy to read - Efficiency - No horizontal scrolls - Easy to read - Easy to read
Search	- Simple search box - Auto suggest	- Easy to search info - Efficiency
Filter	- One click to filter the data - Use of sliders and checkboxes. Less typing.	- Less typing
Action	- Inline editing - Overlay instead of pop-up window	- Efficiency - Perform task within context
Reports	- Interactive chart	- Easy to understand trend - Drill down to details

On the same page, viewing account performance and the comparison to market indexes, tabular presentation instead of graphic chart is used. It is not clear why one part of the site uses interactive chart and another part displays the information in tabular format.

5 Summary and Discussions

Web applications that retrieve data from the server asynchronously in the background and allow users to view information display and interact with the tool at the same time have significantly improved the interaction efficiency. Users can view refreshed information without the interruption of what they are doing on the same screen. They can view information faster in searching and filtering information, taking actions and getting feedback from the system. The immediate feedback from the system also gives users assurance about the actions they have taken. They can fix the error quickly if it occurs.

Rich interaction designs such as inline editing, overlays instead of pop-ups, interactive graphics, and interactive UI components such as slider bars not only have improved the efficiency of interacting with the system but also made it easier for users to complete the tasks. When inline edit feature or an overlay is presented with background information visible, users are kept in the context when performing the task. Rich presentation media such as interactive graphics have made it easier for user to view information. It also enables users to view information over time or across categories and drill information quickly.

Information design has been enhanced in these surveyed personal finance tools in several ways: 1) The amount of information is reduced and more white space is presented. 2) New information related to the current task is displayed. 3) Larger fonts and font variation are used to make it easier for users to read. 4) Contextual help for technical terminologies are provided.

Many personal finance tools have innovative designs in wizards that guide users in complex tasks such as setting up a financial goal, planning for retirement and filing tax returns. Overall, the impact of rich interaction designs and technologies is positive.

Among the personal finance tools, investment applications are the most complex tools that need users to have deeper knowledge of the domain to understand the information presented and take actions. To Kevin and Jane, the average investors who do not have a lot of time and experience in investments but need the vehicle to help them achieve financial goals, the information and interactions must be simple to understand. For example, when an interactive chart is used to show investment performance, it is important for the user to understand what this chart is used for, the terminologies used on the chart, how the values are calculated and what the message the chart sends to the user. The challenge is to make an appropriate design decision when interactive charts and graphics should be applied to the information and how much interactivity should be added.

The biggest challenge to user experience professionals is how to take a complex system, create or apply design patterns that are appropriate to support user goals and task, and work with the development team to implement the designs. This challenge requires user experience professionals to have deep understanding of the user goals and tasks, apply proven design patterns and make decisions when to create new patterns. User experience professionals have to work with the development team very closely to implement the proposed solutions.

References

1. Wikipedia, `http://en.wikipedia.org/wiki/Personal_finance`
2. Garrett, James, J.: The Elements of User Experience: User-Centered Design for the Web and Beyond, 2nd edn. New Riders Press, Indianapolis (2010)
3. Berry, Dick: The user experience – The iceberg analogy of usability(2000), `http://www.ibm.com/developerworks/library/w-berry/#h2`
4. Morville, Peter: User Experience Design (2004), `http://semanticstudios.com/publications/semantics/000029.php`
5. Cerejo, Lyndon: Creating a Hierarchy of User-Experience Needs (2001), `http://www.clickz.com/clickz/column/1696253/creating-hierarchy-user-experience-needs`
6. Owen, R.J.: The Kano Model and the Importance of User Experience (2009), `http://insideria.com/2009/10/the-kano-model-and-the-importa.html`
7. Spool, Jared, M.: Understanding the Kano Model – A Tool for Sophisticated Designers (2011), `http://www.uie.com/articles/kano_model/`

Experimental Investigation of Misuse and Disuse in Using Automation System

Akihiro Maehigashi[1], Kazuhisa Miwa[1], Hitoshi Terai[1], Kazuaki Kojima[2], Junya Morita[3], and Yugo Hayashi[4]

[1] Graduate School of Information Science, Nagoya University, Japan
[2] Faculty of Human Sciences, Waseda University, Japan
[3] School of Knowledge Science, Japan Advanced Institute of Science and Technology, Japan
[4] College of Information Science and Engineering, Ritsumeikan University, Japan
mhigashi@cog.human.nagoya-u.ac.jp,
{miwa,terai}@is.nagoya-u.ac.jp, koj@aoni.waseda.jp,
j-morita@jaist.ac.jp, yhayashi@fc.ritsumei.ac.jp

Abstract. In this study, we experimentally investigated human use of automation systems and the selection strategies of such usage. We used two different types of tracking tasks. As a result, we found that the participants neither tended to misuse nor disuse the automation system. Also, we confirmed that they tended to select to use the automation system depending on their manual performance rather than the system performance. Moreover, we found that there is a relationship between the tendency to use the automation system and the selection strategy.

Keywords: Human-automation system interaction, Misuse, Disuse.

1 Introduction

Many studies have investigated human use of automation systems. It has been pointed out that there are many failures in the use of automation systems, and these failures are defined as misuse and disuse [1]. Misuse is overreliance on automation systems, and disuse is underutilization of systems. Some studies of these tendencies show that people have a bias to overtrust automation systems (automation bias), and tend to misuse them [2][3][4]. Others demonstrate that people have a bias to mistrust automation systems (manual bias), and tend to disuse them [5][6][7][8]. In the current study, we investigate human use of automation systems based on an innovative performance-based analysis. In such a context, we define misuse as using automation systems when manual performance is superior to the system performance. On the other hand, we define disuse as conducting manual operation when the system performance is superior to the manual performance. In this study, on the basis of these definitions, we investigate whether human use of automation systems is inclined to be misuse or disuse.

Moreover, in this study, we investigate human selection strategies for the automation usage. In particular, we define system-based selection as automation usage depending more on system performance than on manual performance. In

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 384–393, 2011.

contrast, we define manual-based selection as automation usage depending more on manual performance than on system performance. In this study, on the basis of these definitions, we investigate whether human selection strategy for automation usage is based on system- or manual-based selection.

2 Experimental Paradigm

2.1 Experimental Task

We used two tracking tasks (line and road tasks) as experimental tasks (Fig. 1). In the line task, the participants track a line that scrolls downward past a circle vehicle. When the circle vehicle veers off the line, the performance score is reduced as operational error. On the other hand, in the road task, the participants track a road that scrolls downward past a dot vehicle. When the dot vehicle hits the edge of the road, the performance score is reduced as operational error. The participants were allowed to switch to either the auto mode (operation completely performed by the program) or the manual mode (operation performed by participants using left and right arrow keys) by pressing a selector on the keyboard. In these tasks, we manipulated the automation capability (Ca) and the manual capability (Cm) with five levels: 30, 40, 50, 60, and 70. Technically, each value indicates the success rate of the operation command. Therefore, the higher the values of Ca or Cm are, the more controllable the circle or the dot vehicle is. In contrast, the lower the values of Ca or Cm are, the less controllable the circle or dot vehicle is.

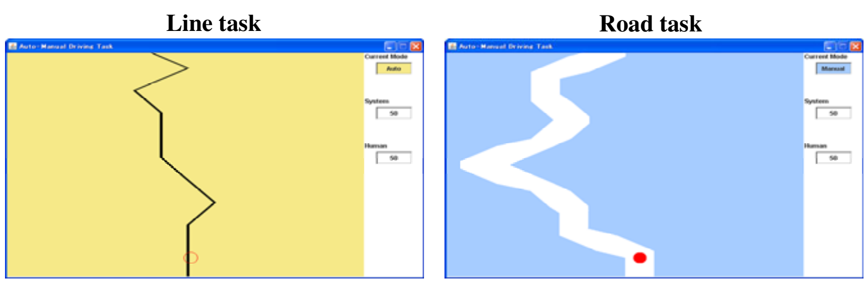


Fig. 1. Experimental Task

2.2 Research Questions

In this study, we investigate three Research Questions described as follows.

RQ 1: Is human use of automation systems inclined to be misuse or disuse?
RQ 2: Is human selection strategy of the automation usage based on either system- or manual-based selection?
RQ 3: Is there any relationship between the tendency to use automation systems and the selection strategy of the automation usage?

2.3 Hypotheses

To investigate the research questions, we conduct an innovative performance-based analysis using logistic regression. First, the average percentage of using the auto mode is recorded at each combination of Ca (5 levels) × Cm (5 levels). Second, we fit the logistic curve to the percentages at the 25 data points. The predicted percentage of using the auto mode is described as follows.

$$\text{Percentage of using auto mode} = 100 * 1/(1 + e^{-(\beta_0 + \beta_1 \text{Ca} + \beta_2 \text{Cm})}). \quad (1)$$

Hypotheses about RQ 1. The point where the values of Ca and Cm are both 50 is the center of the logistic curve. Therefore, we used the percentage of using the auto mode at that point as the representative value indicating the tendency to use the auto mode. Figure 2 shows our hypotheses about RQ 1.

Hypothesis I: Human use of the auto mode is inclined to be misuse. When the percentage of using the auto mode at the representative point exceeds 50%, it means that the auto mode is inclined to be used at the points where the manual performance is superior to the system performance.

Hypothesis II: Human use of the auto mode is inclined to be disuse. When the percentage of using the auto mode at the representative point falls below 50%, it means that the auto mode is inclined to be used at the points where the system performance is superior to the manual performance.

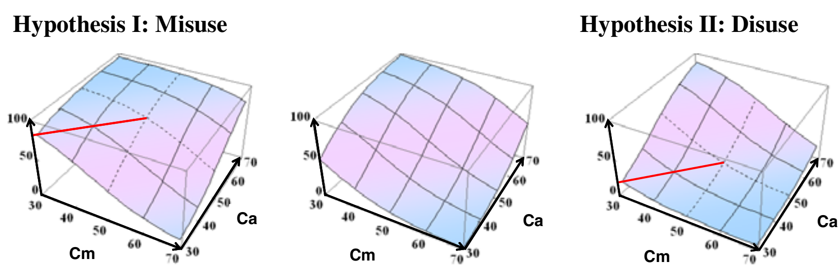
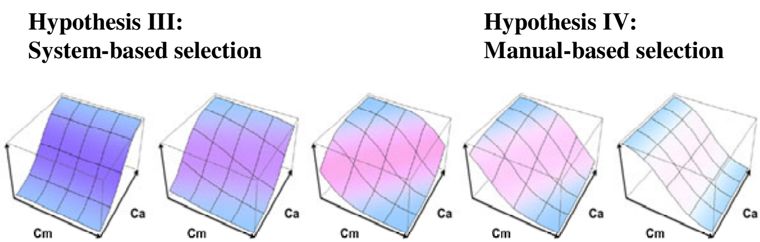


Fig. 2. Hypotheses about RQ 1. The x-axis represents the value of Cm, the y-axis represents the value of Ca, and the z-axis represents the percentage of using the auto mode.

Hypotheses about RQ 2. We use the odds ratios of Ca ($\exp(\beta_1 * 10)$) and Cm ($\exp(\beta_2 * 10)$), calculated from the coefficients of Ca (β_1) and Cm (β_2), and investigate the selection strategy of the automation usage. We compare the increase of using the auto mode with the increase of the value of Ca and the decrease of using the auto mode with the increase of Cm. In particular, we analyze whether the product of the odds ratios of Ca and Cm exceeds 1 or falls below 1. Figure 3 shows the hypotheses of RQ 2.

Hypothesis III: Human selection strategy is based on system-based selection. The product exceeds 1, meaning that the increase of the percentage of using the auto mode with the increase of Ca is greater than the decrease of the percentage of using the auto mode with the increase of Cm. This indicates that the selection of automation usage is made depending more on system performance than on manual performance, and the system-based selection was adopted.

Hypothesis IV: Human selection strategy is based on the manual-based selection. The product falls below 1, meaning that the decrease of the percentage of using the auto mode with the increase of Cm is greater than the increase of the percentage of using the auto mode with the increase of Ca. This indicates that the selection of automation usage is made depending more on the manual performance than on the system performance, and the manual-based selection was adopted.



Product of ORs of Ca and Cm	2	1.5	1	0.75	0.5
OR of Ca	2	2	2	1.5	1
OR of Cm	1	0.75	0.5	0.5	0.5

Fig. 3. Hypotheses about RQ 2. The x-axis represents the value of Cm, the y-axis represents the value of Ca, and the z-axis represents the percentage of using the auto mode.

Hypothesis about RQ 3. We investigate whether people whose styles of using an automation system differ, i.e., misuse-biased users and disuse-biased users, adopt different selection strategies for automation usage.

Hypothesis V: There is a relationship between the human tendency to use the auto mode and the selection strategy of the automation usage.

2.4 Performance-Based Analysis

The procedure is described as follows. First, in each task, we measure the system and manual performances respectively and identify the value of Cm at which the manual performance is equal to the auto performance when the auto capability is Ca.

$$\mathrm{Cm} = A\mathrm{Ca} + B. \tag{2}$$

Next, we replace Ca in the formula (2) with Cm* (the correction value of Cm). By this replacement, when the values of Ca and Cm* are equal, the system and the manual performances are identical.

$$\mathrm{Cm} = A\mathrm{Cm}^{*} + B. \tag{3}$$

Finally, we assign the correction formula (3) to the logistic regression formula (1).

$$\text{Percentage* of using auto mode} = 100 * 1/(1 + \mathrm{e}^{-(\beta_0 + \beta_1 \mathrm{Ca} + \beta_2 (A\mathrm{Cm}^{*} + B))}). \tag{4}$$

In this study, we conduct performance-based analysis, using the correction logistic regression formula (4).

3 Experiment

3.1 Purpose

We investigate Research Questions 1, 2, and 3.

3.2 Method

Participants. Twenty-four university students participated in the experiment. The participants conducted both line and road tasks, and the order of the tasks was counterbalanced among the participants.

Procedure. The participants were required to achieve as high a score as possible in each task, adaptively selecting either the auto or manual mode. In each task, we conducted 25 trials consisting of 5 (Ca: 30, 40, 50, 60, 70) × 5 (Cm: 30, 40, 50, 60, 70). Each trial lasted for 40 seconds. When one trial ended and the other trial began, the display showed “the capabilities changed” on the center of the screen. At the same time, the number of completed trials among the 25 trials was shown. Before conducting each task, the participants performed two training trials for 40 seconds each as practice for switching between the auto and manual modes. In the first training trial, the value of Ca was set to 70 and that of Cm was set to 30. Also, in the second training trial, the value of Ca was set to 30 and that of Cm was set to 70. Throughout the experiment, the values of Ca and Cm were not displayed on the screen: therefore, the participants were not informed of the values.

3.3 Result

First, in each line and road task, the average percentage of using the auto mode was recorded at each combination of Ca (5 levels) × Cm (5 levels). Second, we fitted the logistic curve to the 25 data points. The predicted percentages of using the auto mode are as follows.

- The logistic regression formula in the line task

$$\text{Percentage of using auto mode} = 100 * 1/(1 + \mathrm{e}^{-(0.505 + 0.042\mathrm{Ca} - 0.046\mathrm{Cm})}). \tag{5}$$

- The logistic regression formula in the road task

$$\text{Percentage of using auto mode} = 100 * 1/(1 + e^{-(1.317+0.022Ca-0.044Cm)}). \quad (6)$$

The Hosmer-Lemeshow test was applied to assess goodness of fit of the predicted curves. As a result, the test was not significant in the line (p=.89) and road (p=.97) tasks, indicating that the logistic curves described the data well. Figure 4 shows the predicted curve in each task.

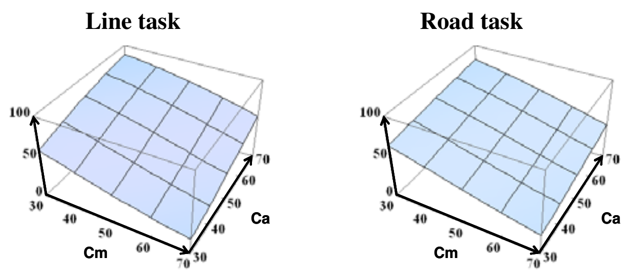


Fig. 4. Predicted Curve of Average Percentage of Using Auto Mode. The x-axis represents the value of Cm, the y-axis represents the value of Ca, and the z-axis represents the percentage of using the auto mode.

In each task, first, we measured the system and manual performances by conducting another experiment in which each task was performed by only either manual or system operation. Next, we calculated the correction formula as follows.

- The correction formula in the line task

$$Cm = 1.177Cm^* - 2.626. \quad (7)$$

- The correction formula in the road task

$$Cm = 1.086Cm^* + 1.897. \quad (8)$$

Finally, we assigned each correction formula to each logistic regression formula (5) and (6), and calculated the correction logistic regression formula as follows.

- The correction logistic regression formula in the line task

$$\text{Percentage* of using auto mode} = 100 * 1/(1 + e^{-(0.505+0.042Ca-0.046(1.177Cm^*-2.626))}). \quad (9)$$

- The correction logistic regression formula in the road task

$$\text{Percentage* of using auto mode} = 100 * 1/(1 + e^{-(1.317+0.022\text{Ca}-0.044(1.086\text{Cm}*+1.897))}). \tag{10}$$

Table 1 shows the percentage of using the auto mode at the representative point, the odds ratios of Ca and Cm, and the product of the odds ratios of Ca and Cm. In Table 1, the percentage of using the auto mode and the odds ratios of Ca and Cm were calculated from the logistic regression formulae, (5) and (6). The percentage* of using the auto mode and the odds ratio of Cm* were calculated from the correction logistic regression formulae, (9) and (10).

Table 1. Percentage of Using Auto Mode at Representative Point, Odds Ratios of Ca and Cm, and Product of Odds Ratios of Ca and Cm

Task	Percentage of using auto mode	Percentage* of using auto mode	OR of Ca	OR of Cm	Product of ORs of Ca and Cm	OR of Cm*	Product of ORs of Ca and Cm*
Line	57.245	50.055	1.522	0.629	0.958	0.579	0.882
Road	54.603	47.652	1.248	0.638	0.797	0.614	0.767

Investigation of RQ 1. Table 1 shows that the percentage* of using the auto mode settled at around 50%. This confirmed that the participants were inclined neither to misuse nor disuse the auto mode. Therefore, the hypotheses I and II were rejected.

Investigation of RQ 2. Table 1 shows that the product of the odds ratios of Ca and Cm* fell below 1. We confirmed that the selection of the automation usage was made based on manual-based selection in both tasks. The participants depended 1.133 (=1/0.882) and 1.303 (=1/0.767) times more on the manual performance than on the system performance in the line and road tasks respectively. Therefore, hypothesis IV was supported.

Investigation of RQ 3. To investigate the relationship between the tendency to use automation systems and the selection strategy of the automation usage, we grouped the participants into two groups: misuse-biased and disuse-biased. In each task, the participants whose average percentage of using the auto mode was higher than the median average were grouped into the misuse-biased group; and the other half of the participants became the disuse-biased group.

In each misuse- and disuse-biased group engaging in each line and road task, the average percentage of using the auto mode was recorded at each combination of Ca (5 levels) × Cm (5 levels). Second, we fitted the logistic curve to the 25 data points. The predicted percentages of using the auto mode are as follows.

- Misuse-biased group in the line task

$$\text{Percentage of using auto mode} = 100 * 1/(1 + e^{-(0.249+0.047\text{Ca}-0.034\text{Cm})}). \tag{11}$$

- Disuse-biased group in the line task

$$\text{Percentage of using auto mode} = 100*1/(1+e^{-(0.522+0.045Ca-0.062Cm)}) . \tag{12}$$

- Misuse-biased group in the road task

$$\text{Percentage of using auto mode} = 100*1/(1+e^{-(1.696+0.031Ca-0.047Cm)}). \tag{13}$$

- Disuse-biased group in the road task

$$\text{Percentage of using auto mode} = 100*1/(1+e^{-(1.080+0.017Ca-0.048Cm)}) . \tag{14}$$

The Hosmer-Lemeshow test was not significant in the misuse- (p=.94) and disuse- (p=.78) biased groups in the line task, or in the misuse- (p=.95) and disuse- (p=.93) biased groups in the road task, indicating that the logistic curves described the data well.

Table 2 shows the percentage of using the auto mode at the representative point, the odds ratios of Ca and Cm, and the product of the odds ratios of Ca and Cm. In Table 2, the percentage of using the auto mode and the odds ratios of Ca and Cm were calculated from the logistic regression formulae (11)-(14). The percentage* of using the auto mode and the odds ratio of Cm* were calculated from the correction logistic regression formulae in which the correction formula (7) was assigned to the logistic formulae (11) and (12) for the line task, and the correction formula (8) was assigned to the logistic formulae (13) and (14) for the road task.

Table 2. Percentage of Using Auto Mode at Representative Point, Odds Ratios of Ca and Cm, and Product of Odds Ratios of Ca and Cm

Task	Group	Percentage of using auto mode	Percentage* of using auto mode	OR of Ca	OR of Cm	Product of ORs of Ca and Cm	OR of Cm*	Product of ORs of Ca and Cm*
Line	Misuse	71.054	66.482	1.600	0.711	1.138	0.669	1.071
	Disuse	42.192	33.070	1.578	0.535	0.845	0.479	0.757
Road	Misuse	71.490	65.156	1.372	0.623	0.856	0.598	0.821
	Disuse	38.319	31.499	1.189	0.616	0.732	0.590	0.702

In each task, we compared the products of the odds ratios of Ca and Cm* in the misuse- and disuse-biased groups. As a result, the product was lower in the disuse-biased group than in the misuse-biased group. This result showed that the participants in the disuse-biased group made the selection of the automation usage depending more on manual performance than those in the misuse-biased group did. We confirmed that there is a relationship between the tendency to use automation systems and the selection strategy of the automation usage. Therefore, hypothesis V was supported.

4 General Discussion

4.1 Human Tendency to Use Automation Systems

Some previous studies showed that people have a bias to overtrust automation systems (automation bias), and tend to misuse them [2][3][4]. In these studies, the automation systems performed one part of the multiple tasks, almost always worked normally, and rarely caused system failures. In these experiments, the participants misused the automation systems because they could not monitor system errors properly. In our experiment, the participants were required to select to use either the system or manual operation in a single task. Additionally, the performances of system and manual operations dynamically changed. From these differences from the previous studies, our participants continuously monitored the system performance, and did not tend to misuse the automation system.

On the other hand, other preceding studies showed that people have a bias to mistrust automation systems (manual bias), and tend to disuse them [5][6][7][8]. In these previous studies, the participants were required to select to use either the system or manual operation in a single task. In the initial block of trials, the participants and the automation system concurrently performed the task, and they were informed of the performances of their manual and system operations after each trial. In the second block of trials, the participants were required to select to use either the system or manual operation based on the experiences for training in the initial block. These studies showed that the participants disused the high-performance systems in the second block because of the lack of capability for objective evaluation [5], human competition with automation systems [6], and strong human sensitivity to system errors [7][8]. In our experiment, to maximize the task performance, the participants were required to select to use either the system or manual operation, continuously evaluating the performance in each operation. From these differences from the previous studies, our participants could objectively evaluate the system and manual performances and did not tend to disuse the system.

4.2 Selection Strategy

In this study, we investigated the selection strategy of the automation usage. As a result, we found that the selection of the automation usage was made based on manual performance and manual-based selection was confirmed. In our experiments, the participants needed to evaluate the performances of the system and manual operations in order to maximize the task performance. However, they might not be able to consider such evaluations from both viewpoints evenly because of the limitation of human cognitive capacities. This limitation led the participants to select automation usage based on a single viewpoint. Moreover, the previous studies about situation awareness showed that people have superior abilities in active monitoring (monitoring situations while manually conducting a task) compared to passive monitoring (monitoring situations while observing an automation system's operation) [9][10]. It is assumed that our participants adopted manual-based selection because of such superiority of active monitoring.

Furthermore, we found that there is a relationship between the tendency to use automation systems and the selection strategy of the automation usage. There is a possibility that the participants in the disuse-biased group were inferior in abilities to evaluate system performance than those in the misuse-biased group.

References

1. Parasuraman, R., Riley, V.: Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors 39(2), 230–253 (1997)
2. Bahner, J.E., Hüper, A.-D., Manzey, D.: Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. International Journal of Human-Computer Studies 66(9), 688–699 (2008)
3. Singh, I.L., Molloy, R., Parasuraman, R.: Individual Differences in Monitoring Failures of Automation. The Journal of General Psychology 120(3), 357–373 (1993)
4. Skita, L.J., Moiser, K., Burdick, M.D.: Accountability and automation bias. International Journal of Human-Computer Studies 52(4), 701–717 (2000)
5. Beck, H.P., Dzindolet, M.T., Pierce, L.G.: Automation Usage Decisions: Controlling Intent and Appraisal Errors in a Target Detection Task. Human Factors 49(3), 429–437 (2007)
6. Beck, H.P., McKinney, J.B., Dzindolet, M.T., Pierce, L.G.: Effects of Human-Machine Competition on Intent Errors in a Target Detection Task. Human Factors 51(4), 477–486 (2009)
7. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The role of trust in automation reliance. International Journal of Human-Computer Studies 58(6), 697–718 (2003)
8. Dzindolet, M.T., Pierce, L.G., Beck, H.P., Dawe, L.A.: The Perceived Utility of Human and Automated Aids in a Visual Detection Task. Human Factors 44(1), 79–94 (2002)
9. Endsley, M.R., Kiris, E.O.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. Human Factors 37(2), 381–394 (1995)
10. Metzger, U., Parasuraman, R.: The Role of the Air Traffic Controller in Future Air Traffic Management: An Empirical Study of Active Control versus Passive Monitoring. Human Factors 43(4), 519–528 (2001)

Validating Video Analytics in Mission Critical Applications

Stephen J. Mitchell[1], Sukhpreet Gill[1], Steve Loveless[1], and Brent Auernheimer[2]

[1] Pelco, 2500 Pelco Way, Clovis CA 93612-5699 USA
[2] Computer Science, California State University, Fresno CA 93740 USA
{smitchell,sukhpreet.gill,sloveless}@pelco.com,
brent@csufresno.edu

Abstract. Video Analytics (VA) automates and aids intelligent decision making in video surveillance applications. A video security system combines sophisticated video analytics with the oversight and judgment of human operators to save lives and critical resources. The complexity of the user experience and the range of deployments of VA make testing particularly challenging. Trade-offs between functional tests and focused usability tests are used to build a case for trustworthiness. This paper describes a method for validating video analytics for use in mission critical applications.

1 Introduction

Users depend on Video Analytics (VA) functionality in mission and safety critical situations like airports, banks, museums and casinos. Users need assurance that they can interact with VA reliably and efficiently during critical and non-critical times. However, VA functionality and interfaces can be daunting and users overwhelmed with the complexity of configuration, use, and verification. How can developers test a VA system to ensure users will correctly configure the product for their environment and be able to use this product in critical moments?

1.1 Video Security

The video security industry provides products to assure customers that they, and their valuables, are safe. Traditionally this meant that security personnel monitor a number of cameras that, in turn, monitors the valuables and records events. Security personnel observe multiple camera monitors simultaneously, which can lead to fatigue and errors. The hope and goal of VA is to augment the performance of security personnel [4].

1.2 Video Analytics

The purpose of VA is to "describe the real time use of computer vision in a security environment to monitor the CCTV camera feeds and assist the guard in his or her decision making process" [8]. VA provides role-based features to monitor digital video and communicate to security personnel when designated events occur. VA typically includes functions such as Adaptive Motion, Abandoned Object, and Object Counting:

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 394–402, 2011.

Adaptive Motion detects and tracks objects that enter a scene, triggers an alarm when the objects enter or touch a user-defined zone, and monitors the objects until they exit the scene. Typical installations are banks, casinos, homeland security, the retail industry and public transportation.

Abandoned Object detects objects placed within a predefined zone and triggers an alarm if the object remains in the zone too long . Security personnel use this algorithm for detecting suspicious objects in a scene. Typical installations for this function are airports and public areas.

Object Counting calculates the number of objects that enter a defined zone or cross a virtual trip wire and then triggers an alarm once the count reaches a predefined number. Typical installations for this behavior include counting people (or vehicles) at an exit/entry door (or parking lot).

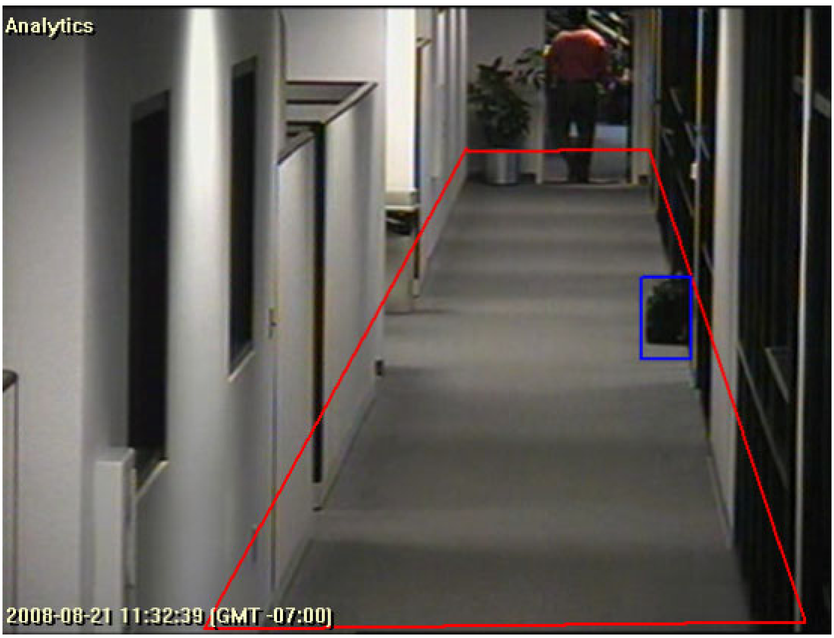


Fig. 1. An abandoned object

2 Validating Video Analytics: Two Techniques

We consider two techniques to build customer trust and assure users of correct VA operation: functional testing (component, then system level) and focused usability testing of the system in context. In order to provide high performance and a rich user experience, we recognize three elements of context: users, equipment, and environment.

After developers implement features they believe users need, functionality is tested at the component and system level. Then, after validating functionality of each component, and then components together as a system, testers need to validate the interaction of users and the system in varied environments. Focused usability testing fulfills this need.

Focused usability testing is inspired by ethnographic approaches to usability testing. Hollan et al. describe a distributed cognition approach to the interaction of users and complex systems saying that "Distributed cognition ... is specifically tailored to understanding interactions among people and technology" [2]. However, Don Norman cautions against unfocused approaches to this type of testing, saying that explorative ethnography helps in interpreting human behavior but cannot be applied to the real-world due to its dispersed nature [10]. We believe that focused usability testing not only helps determine if users can effectively use the system, but also ensures the system responds to diverse needs of users. The goal is to provide a rich user experience.

3 Problems Validating Video Analytics

While VA developers intend for their applications to be comprehensive and useful, we show later in this paper that users can find them complex and difficult to configure. Also, even after users configure the system to suit their environment, they can be misled and their expectations be faulty. Sometimes the VA application does not facilitate users developing their own assurance that they have successfully configured the functions for their purposes. Our experience leads us to believe that a combination of user interface improvements as well as user preparation is missing. Developers and testers must consider this throughout the software development process.

3.1 Functionality Does Not Suffice

Functionality for VAs is, by its nature, tied to a real-world scene. VA cannot operate without a scene. The process of validating functionality consequently leads to examining how customers configure the system for their environment. When VA requires configuring, the software experts (usually the software developers) put the performance of the product in to the hands of the novice (the software user).

3.2 Chaos in Context of Use

Context of use can be defined as "users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used" [3] (see Fig. 2). The context of use for VA systems varies greatly due to the chaotic nature of the environment (scene).

While VA developers can document required general environmental conditions for operation, it is difficult to be precise, concise, and yet capture the range of real-world environments. Users are therefore required to interpret what real-world environmental conditions apply to them, and then ensure that they control their environment as much as possible (such as camera placement and choice of scenes). This minimizes the variability of the context of use. After making proper adjustments for their environment, users then have to configure their VA system to capture the data that they need.

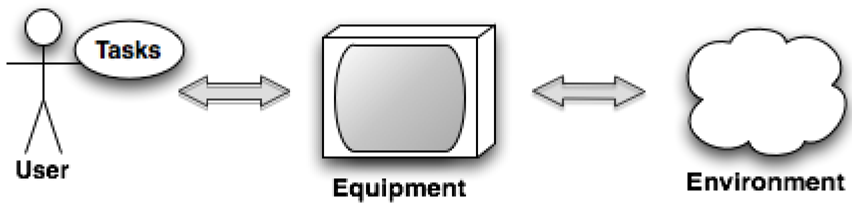


Fig. 2. Context of Use

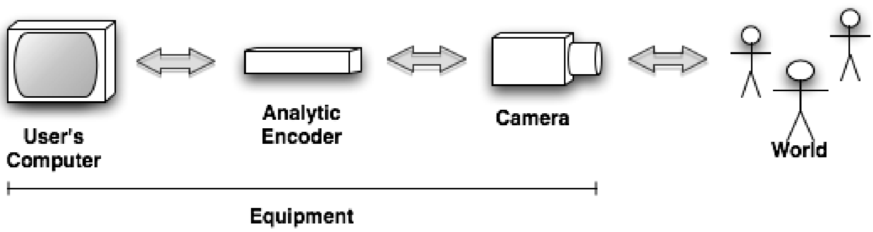


Fig. 3. An Environment for Video Analytics

A characteristic of video systems is they provide users an interface to a collection of equipment, which itself interfaces to the real world (see Fig. 3). Events in the real world may differ day to day, complicating the context of use. Worst case, these real-world scenarios are chaotic.

4 Applying Standards to Context of Use

We recognize three elements of context of use which play important roles in the performance of VA: users, equipment, and environment.

4.1 User: Training and Experience

Don Norman says "People form mental models through experience, training and instruction. The mental model of a device is formed largely by interpreting its perceived actions and its visible structure" [9]. User interfaces, like those for VA, can be complex, creating a potential mismatch between the developers' implementation model and the users' mental models. Training and experience help form a user's initial mental model.

We provided training to users in the form of face-to-face training as well as through written instructions and manuals. In the former method, trainers configure VA on a representative scenario and then give the user control to configure the system. Training ends with an open-ended session where users voice questions and concerns. Manuals are also provided.

We also consider constructivism when assessing the performance of VA users. Martin Dougiamas defines constructivism as "building on knowledge known by the student" [7]. That is, users learn not only by instruction but also by experimentation. Users tweak parameters to make their own discoveries and conclusions.

But because users' environments vary greatly, users cannot solely depend on training to prepare them for achieving their goals for VA. Users must have an interface that directs them and provides feedback so they gain confidence in the system.

4.2 Equipment: Usability Study

When conducting our VA usability study we followed guidelines in NISTIR 7432 [5]. This standard provides information required to conduct usability tests and serves as a test protocol. After recruiting users, and attaining test scripts and a test system, moderators carry out the usability test. Once users have executed the tasks, the moderator gathers data from the users through a questionnaire and one-on-one interviews.

4.3 Environment: i-LIDS

In order to control context-of-use variables, we constrain the test environment. The UK government's Home Office Scientific Development Branch offers standardized sets of videos, "i-LIDS", for this purpose [1]. i-LIDS provides the precise conditions of the environment the video represents as well as how the VA algorithm under test should behave. The description of how the VA should behave is a performance standard for each VA function. Our usability studies used i-LIDS videos representing low chaotic environments so that users in the studies will have a higher potential for positive results.

4.4 Standards

IS0 9241-11 emphasizes that "Quality In Use" (QIU) depends on context of use and that the level of QIU achieved depends on the specific circumstances in which a product is used [3]. Also, ISO 9126 states that in order to ensure customer satisfaction, quality must be established at each phase of the development of a product [3]. "Internal" quality characteristics are tested when defining the product. For example, requirements and specifications fulfill metrics that were a result of considering desired quality characteristics. Testers then apply "external" quality characteristics to the product to ensure that developers implemented requirements and specifications properly [4].

In order to better assess if customers will be satisfied, we applied ISO 9126 QIU quality characteristics. QIU testing is similar to usability testing in that testers observe users with the product to determine if goals for the product are met [3]. When testers measure QIU, they are effectively measuring how well internal and external quality characteristics were applied for the end user [4]. ISO defines four software QIU characteristics within a specified context of use:

- *Effectiveness*: enabling users to achieve specified goals with accuracy and completeness.
- *Productivity*: enabling users to expend appropriate amounts of resources in relation to effectiveness.
- *Safety:* achieving acceptable levels of risk of harm to people, business, software, property or the environment.
- *Satisfaction:* satisfying users.

5 Applying "Quality in Use"

Because the performance of VA depends on the ability of the user to configure the analytics, we emphasize that user effectiveness is essentially tied to overall success.

5.1 Our Quality in Use Test

QIU tests how well users are able to configure the system so that it will detect an event defined by the VA function. We measured the performance of users and VA functions as a single system in terms of the QIU characteristics listed in Section 4. Metrics were used to compare the number of actual events in a video to how well the VA algorithm detected the events given a user's configuration. The VA industry

Table 1. Metrics Used Applying QIU (based on [4])

Characteristic	Metric	Measurement	Interpretation
Effectiveness	1. User/System Effectiveness 2. Error frequency	1. X = A/B 2. Y = C/B A = number of true positives B = number of ground truths C = number of false positives	1. 0 <= X <=1 The higher the better 2. 0 <= Y <= 1 The lower the better.
Productivity	1. Task time 2. Economic productivity	1. X = mean time taken by users to complete assigned tasks 2. Y = A/B A = task effectiveness B = total task cost	1. 0 minutes <= X <= 15 minutes The lower the better 2. 0<=Y The higher the better
Safety	Safety of people affected by the system	X = 1-A / B A = number of false negatives B = number of ground truths	0 <= X <= 1 Closer to 1 is safer.
Satisfaction	1. Satisfaction scale 2. Satisfaction questionnaire	1. X = Rating on Likert scale [4] 2. Y = A/B A = positive comments B = total comments	1. 0 <= X <=5 5 is completely satisfied. 2. 0 <= Y <= 1 The higher the better.

refers to actual events in a video as "ground truths", correctly identified events are referred to as "true positives", incorrectly identified events are referred to as "false positives" (i.e., if reported event was not an actual event), incorrectly unreported events are referred to as "false negatives" [14].

In order to verify the VA interface is simple to use, and to discover errors and areas for improvement, two rounds of usability studies were conducted.

Round One. We tested both expert and novice users. We expected expert users to have higher values of effectiveness and productivity than novice users, but after conducting the first round of usability study this was not the case. Results indicated that although the productivity of expert users was double the productivity of novice users, there was not much difference between the effectiveness of expert and novice users.

We explained the higher productivity of expert users by the novices' learning curve [6]. The possible explanation for lower effectiveness of expert users relative to novice users might be explained by the Hawthorne effect [5]. We observed that novices were more concentrated and performed each task carefully whereas experts overlooked some of the simple configuration parameters that play an important role in VA. Novices may have the perception that their performance was being judged and hence performed conscientiously, whereas the level of attention given by experts was low and therefore their performance was poor. We also speculate a reason for low performance was due to the use of more chaotic video scenes for experts and less chaotic videos for novices.

When analyzing recorded screens and videos of users in round one of usability testing, we discovered a common set of mistakes committed by most users that resulted in the decrease of effectiveness of the system. We concluded that specific training might address these errors and make a difference in performance. We conducted another round of usability testing using 12 novice users, where six were "normally" trained and another six were "specially" trained.

Round Two. Because we found interface and functionality problems in the Object Counting algorithm, we decided to test Adaptive Motion (first) and Abandoned Object (second) for this round of testing. We also decided to more tightly control our test environments (which are part of the context of use for the test), thus widening the gap between a "specially trained" group and the "regularly trained" group. Regularly trained users were provided basic instruction to acquainting them with the look, feel, and operation of the VA system. They were instructed about the parameters in the setup that could affect the functionality of the VA functions. Specially trained users not only received "regular" training, but also were given specific knowledge about tweaking each parameter in the configuration and the effect on VA performance.

5.2 Quality in Use Results

Again, we measured the performance of users and analytics together as a system using metrics of QIU characteristics.

Effectiveness. The results indicated that effectiveness of specially trained users was 40 percent higher than that of trained users for both Adaptive Motion and Abandoned Object algorithms. Error frequency for specially trained users was 2 percent less than that of trained users. This indicates that specially trained users are more apt to properly respond to real events.

Productivity. Productivity of specially trained users was 27 percent greater than that of normally trained users. Also, when we compared productivity for Adaptive Motion and Abandoned Object, it was higher for Abandoned Object for both trained and specially trained users as they had more experience while conducting tasks on the Abandoned Object algorithm (all 12 users performed tests on Adaptive Motion first and then on Abandoned Object). This confirms that productivity increases with experience.

Safety. False negative counts could represent occurrences where something bad might have happened but users weren't alerted. When comparing results from our QIU testing, specially trained users' environments were 51 percent safer than the regularly trained users' configurations.

Satisfaction. When we analyzed satisfaction, we found that trained users were four percent less satisfied than specially trained users. The satisfaction questionnaire indicated that specially trained users were 14 percent more satisfied than trained users. Not surprisingly, satisfaction goes hand in hand with performance: users are more satisfied if they can use the system effectively.

6 Conclusion

As with any software product, testers must thoroughly validate VA before product release. In order to ensure product readiness for a variety of users in diverse environments, QIU testing for VA shows that functional testing is not sufficient to show product readiness. QIU shows that VA developers should train their users in a controlled environment so that users learn how the software interacts with various environments. Unprepared users will not have confidence in a newly acquired VA system. Also, if we only train users on how to use the equipment, they stand a higher risk of using the system poorly. Users must be trained on their entire context of use. If users understand their environments in relation to their equipment, they mitigate variability (similar to our use of i-LIDS in our QIU testing). If customers are deploying VA in a mission or safety critical environment, such as an airport or casino, lack of training could mean missed opportunity for saving lives or resources.

We've shown the importance of end-user training for the successful VA deployment. This research will help guide the development of user training as well as VA with affordances to enable successful operation

Acknowledgements. Pelco is the video security brand of Schneider Electric, the global specialist in energy management. Pelco is fully integrated within the Schneider Electric Buildings Business Unit. The work illustrated in this publication is the result of ongoing product development in the Buildings Business Unit's Video Product Development department and is supported by the Building Business Unit's Solutions and Technology Office. Thanks to Greg Millar, CTO and Sr. VP of Solutions and Technology, and to Ray Still, Director of Systems Software for supporting this work.

References

1. i-LIDS Team: Imagery Library for Intelligent Detection Systems (i-LIDS): A Standard for Testing Video Based Detection Systems. In: Proceedings 2006 40th Annual IEEE International on Carnahan Conferences Security Technology, pp. 75–80. IEEE Press, Los Alamitos (2006), doi:10.1109/CCST.2006.313432
2. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction 7, 174–196 (2000)
3. ISO/IC 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. International Organization for Standardization (1998)
4. ISO/IEC 9126. Software engineering - Product quality. International Organization for Standardization (2002)
5. NISTIR 7432. Common Industry Specification for Usability - Requirements Annex C. National Institute of Standards and Technology. IUSR: CISU-R v0.90, pp. 22–24 (June 2007)
6. Kitchenham, B.A., Sadler, C.: Evaluating Software Engineering Methods and Tools Part 4: The influence of human factors. ACM SIGSOFT Software Engineering Notes 21, 11–13 (1996)
7. Dougiamas, M.: A Journey Into Constructivism (1998), `http://dougiamas.com/writing/constructivism.html#intro`
8. Video Content Analysis, What Is It and Why Would I Want It? `http://smartcctv.blogspot.com/2006/11/video-content-analysis-what-is-it-and.html`
9. Norman, D.A.: The Design of Everyday Things. MIT Press, Cambridge (2002)
10. Norman, D.A.: Workarounds and hacks: the leading edge of innovation. Interactions 15, 47–48 (2008)
11. Pelco. Endura Abandoned Object Behavior Operation/Configuration manual C3602M (2007)
12. Pelco. Endura Adaptive Motion Behavior Operation/Configuration manual C3600M (2007)
13. Pelco. Endura Object Counting Behavior Operation/Configuration manual C3603M (2007)
14. Yin, F., Makris, D., Velastin, S.: Performance Evaluation of Object Tracking Algorithms. In: Proceedings of the 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. In: PETS 2007 (2007), `http://www.cvg.rdg.ac.uk/PETS2007/data.html`

Proposal of an Office Work Productivity Model Based on Short Pauses in Mental Tasks

Kazune Miyagi, Show Kawano, Hong Zhe Jin, Hiroshi Shimoda, and Hirotake Ishii

Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto, Japan
{miyagi,kawano,hongzhej,shimoda,ishii}@ei.energy.kyoto-u.ac.jp

Abstract. In this study, the authors propose a model which can explain and simulate the change of office work productivity. This model assumes office workers transit a working state and a non-working state with the probabilistic functions based on the variation of brain fatigue. Aiming to discuss the detail of the model, the authors have conducted a subject experiment in which work motivation are controlled. Comparing the results of the experimental results with the result of computer simulation based on the model, it was found that the subjects tend to concentrate the task when the remuneration is provided, and they try to keep their attention to the task and take a long break instead when the task workload is high.

Keywords: Productivity, Working State, Probabilistic Functions, Human Modeling.

1 Introduction

Recently, many office building have been trying energy saving to reduce CO2. Cool Biz, the activity wearing casual style clothes and setting air conditioners at 28 degrees in the summer, is a major activity in Japan. However, we should consider not only wattage per hour but also comfort of office. Recent studies have revealed that improvement of office environment may improve the productivity of office workers [1]. If the office condition takes working time long, there is a risk of increasing total energy consumption.

The authors, therefore, had developed an office work performance test, CPTOP (Cognitive Performance Test for Office Productivity), aiming at establishing an evaluation method of office work productivity [2]. In addition, the authors proposed a circadian lighting system, which promote productivity by adjusting circadian rhythm with high luminance light, and conducted laboratory experiments with CPTOP. As the experimental result, it was found that the results of both CPTOP and simulated office work under the circadian rhythm lighting were improved approximately 4% comparing with the normal lighting [3].

As a result of many studies aiming to reveal relationship of productivity and indoor environment (ventilation, temperature, noise, etc.), the environment which promotes office work productivity becomes elucidated.

However, the mechanism of its change is unexplained and the size of it also depends on tasks and human internal factors. Therefore, the author is trying to create

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 403–412, 2011.

a model which can explain and simulate the change of office work productivity. In this study, as the first step of creating model, the authors have conducted a subject experiment focusing on internal factors. In this experiment, the indoor environment was fixed. Because, the effect size of motivating factor is larger than indoor environment. The motivating factor is suitable for considering about the process of productivity change.

With the experiment result, authors discussed the details of the model by comparing the results of the computer simulation based on the model with the experimental results. If this model is completed by the result of experiment, which the authors are planning to conduct for revealing the effect of indoor environment, we can predict productivity without experiment, and optimize the balance between productivity and energy use by the model.

2 Proposal of the Model

2.1 Target Extent of the Model

There are much type of office work and factors which affect productivity. These factors also affect each other. Because of its complex relationship, it is difficult to make a model considering all factors. In this study, authors limit target extent of the model.

2.2 Factors

Indoor environments are one of the factors which affect work place productivity, and many studies are focusing on them. However, their affect is not only physiological but also psychological. In addition the size of productivity change caused by them is small and its size variable. On the other hand, it is quite unlikely to human inner factors, like motivation, change indoor environments, and their effect size is large. In this study, the proposing model is aiming to explain the process of productivity change. And the type of factor which causes productivity change is not important. In this study, we focus inner factors, especially motivation and mental workload.

2.3 Contents of the Work

There are various kinds of office works, and the human abilities for office work also various. But, works which occupy a considerable amount of working time are mental tasks which have standard routine, for example, paperwork or information processing. These mental tasks which use conscious symbol processing can be explained using with analogy of work of single processor computer. Using this analogy makes it easy for modeling the process of office work and to simulate the office works on computers. Computer simulation would be useful for predict productivity change. Because of it, the type of office work in the model is using conscious symbol processing. On the other hand, the model doesn't consider about more complex tasks, which need advanced creative abilities.

2.4 Working Style

The working style in real office is seem that office workers devote a given time period for their work. The period would be more than 30 minute or several hours. And workers address their works at their own pace In this period. In this study, working style are assumed as above one.

2.5 Time-Series Analysis of Previous Experiment

Aiming to guess the mechanism of work productivity change, the feature of an experimental result was extracted with time series charts. This experiment conducted in 2008[4] was aiming to compare task performance under bad environment (30 degrees, 750lx) with good one (26 degrees, 2500lx). As a result, the performance of receipt checking task in good environment was higher than one in bad. In receipt checking task, there are many 200 paper receipts and the information (Amount of money, Date, *etc*…) corresponding receipts are shown in PC screen. Subjects should check the information error in seven teen minutes.

Figure 1 is one of the experiment results. The axis of ordinate shows the times of checking one receipt. Abscissa axis shows the lapsed time. As shown in fig 1, the mid-level time of checking receipt looks similar in both environments. On the other hand, the numbers of sheets for which the subject took more than twice time longer than mid-level time looks similar in both environments. The numbers clearly looks higher in bad environment than good environment. A receipt checking task is simple task and the time for checking is almost same in every receipt. It means that the time subject takes long for checking is shouldn't be for only checking but including short time pause. For this reason, the authors assumed that the main process of productivity change is these pauses and this pauses play a role of relieving fatigue.

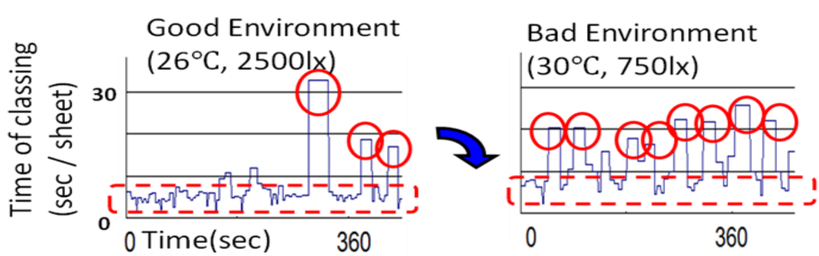


Fig. 1. Time series analysis of previous experimental result

2.6 Framework of the Model

From above analyzing, the authors propose an office work productivity model based on short pause. Figure 2 shows the framework of this model.

1. Office workers are assumed to transit a working state and a non-working state.
2. The transition of these two states is along with the probabilistic functions based on the variation of BF. BF is a hypothetical value which is assumed as a brain fatigure.
3. Office workers are assumed to transit these two states along with the probabilistic functions based on the variation of BF.
4. At working state, the working progress goes off and value of BF increase. On the other hand, at working state, working progress stops and value of BF decrease.
5. The transition probability is determined by human inner factors, indoor environment and so on.

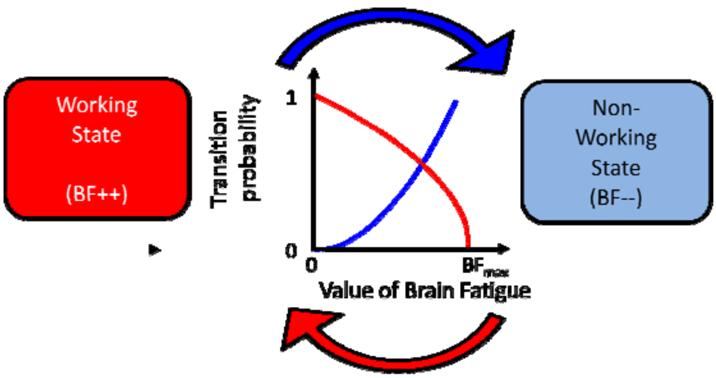


Fig. 2. Framework of the model

2.7 Computer Simulation Based on the Model

This model is created with an analogy of computer and can simulate the change of working state and productivity. Concretely, if two transition probability,P_1(BF) means from working to non-working, P_2(BF) is from non-working to working, and the change speed of BF is determined, model simulate how the works progress and simulate productivity with probability. In this study, the authors used equation (1) to (3) as transition probability and change speed of BF. These equations are determined aiming to explain the change of equations with few parameters.

Parameter a_1 and a_2 means the tendency to transit. If these parameters become higher, the probability becomes higher at same value of BF. Parameter v_1 and v_2 means the change speed of BF.

$$P_1(BF) = \left(\frac{BF}{BF_{max}}\right)^{a_1} \quad \text{(working to non-working)} \qquad \cdots(1)$$

$$P_2(BF) = \left(\frac{1-BF}{BF_{max}}\right)^{a_2} \quad \text{(non-working to working)} \qquad \cdots(2)$$

$$\frac{dBF}{dt} = \begin{cases} v_1 - v_2 \cdot BF \text{ (wroking state)} \\ \quad\;\; v_2 \cdot BF \text{ (non working state)} \end{cases} \qquad \cdots(3)$$

3 Subjective Experiment Controlling Inner Factors

3.1 Purpose

The experiment in 2.2 had problem that the number of subject is few and the experimental condition was complex. The purpose of this subjective experiment was collecting the time-line data of working under motivation and mental work load was controlled. In addition, by comparing experiment result and simulation result based on the model, the model parameter derived on each condition and discuss about this model.

3.2 Experimental Method

In this experiment, the task performance was measured in two motivational conditions and in two mental work load condition and compared. This experiment was conducted on Nov 24 to 29 in 2009. Figure 3 shows a scene of experiment. During this experiment, the environment was fixed. The illumination on the desk was 750lux, temperature was 25 degrees, CO_2 concentration was under 800ppm and noise level was under 52dB. Twelve subjects joined this experiment, who are seven males and five females. The average age was 21.7.

Fig. 3. a scene of experiment

Controlling inner factors. In this experiment, motivation to the work was controlled. In “high motivated” condition, the subject was explained that they could get additional payment according to their task performance. In “low motivated” condition, there was no additional payment. The motivation was checked with a subjective assessment index. In addition, the mental load work to the work was controlled to two conditions by task difficulty. The mental work load was checked with NASA-TLX [XX].

Task. The task used in this experiment was mental addition task. As shown in Figure 4, the first numbers is displayed in window, and when enter key was pressed the second number is displayed. In this task the number of added the first and second number should be inputted. The number shown was three digits in “low mental work load” condition, and four digit in “high mental work load” condition.

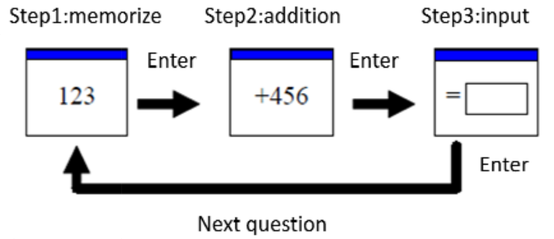


Fig. 4. Mental addition task

Experimental procedure. The experiment was conducted for one day per person. Figure 6 shows the experimental procedure in each day. After the time of explaining and practice of mental addition task, eight set of mental additional task were given to subjects. The time of one set is thirty minutes. After Each task, the questionnaire was conducted. And short breaks and long lunch time break was given among each tasks.

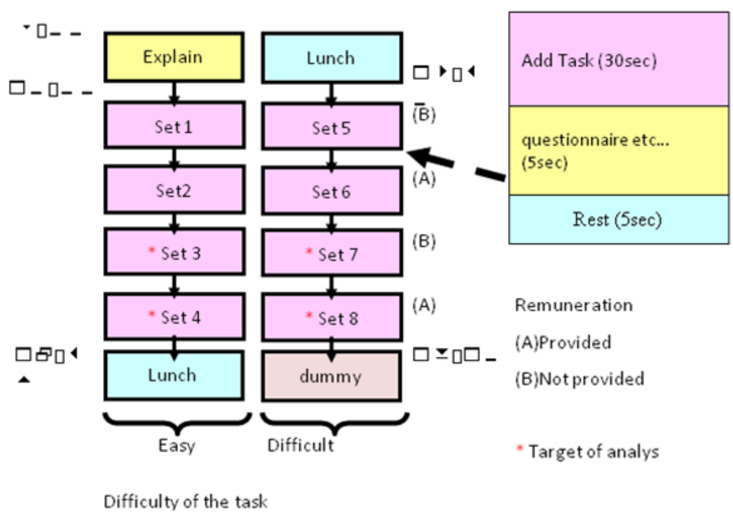


Fig. 5. Experimental procedure in a day

Measured Indexes. The performances of mental additional task were measured as performance indexes. Fatigue questionnaire examines the fatigue states of the subjects in five viewpoints which are sleepiness, discomfort, haze, instability and dullness. It consists of 25 questions and they are answered with five grades. Motivation questionnaire examines the subjective motivation from zero point to a hundred points. NASA-TLX examines the mental work load. From some question, mental work load is measured in one hundred degrees.

3.3 Experimental Results and Discussion

3.3.1 Task Performance and Inner Factors

Figure 6 shows the average of the performance indexes of each task set and the score of subjective motivation and NASA-TLX. As the result, providing additional payment improve their motivation, and the task performance in "high motivated" condition is higher than in "low motivated" condition ($p < 0.001$). And high difficulty of the task improve mental work load as shown in NASA-TLX score when the motivational condition was "high motivated" ($p<0.05$). On the other hand, there is no significant difference when the motivational condition was "low motivated". It may means that subject work easy at their own pace when low motivated, and it works mental work load to be fixed. However, these results don't have a particular meaning. This experiment was planned to get the result in which task performance is clearly changed.

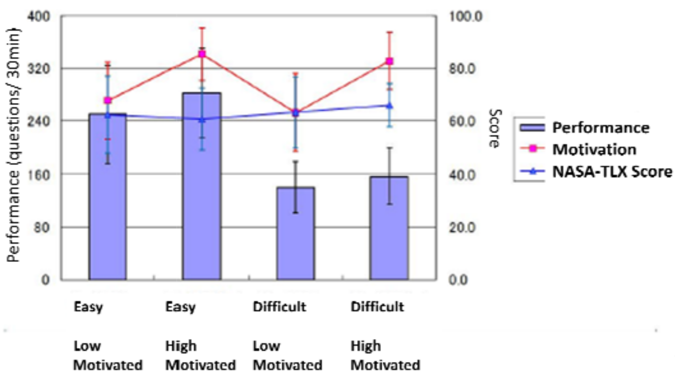


Fig. 6. Performance of mental addition task

3.3.2 Model Parameter

The model parameters in equation (1) to (3) to simulate experimental results have been examined with following method. When the four parameter a1, a2, v1 and v2 determined, a time line data of working is simulated by computer, for example, shown in top of figure 7. From this result, histogram can be made. By optimize the model parameters to minimize the similarity of the form of histogram of simulated result with one of experiment data. The similarity in examined as Error rate. It was defined, as shown in figure 8. By comparing the histogram of answering time, square sum of each variation between simulation result and experiment result was calculated. And, the parameter which makes the error rate smallest was calculated. Maximum likelihood estimation method was used to optimize. The parameters were determined for each task set of each subject.

Table 1 shows the average of the model parameters and result of significant test. And figure 9 shows the form of transition probability in high motivated and low motivated condition. When motivation becomes high, parameter a_1 became higher ($p<0.001$) and a_2 became lower. When mental work load becomes high (the task became difficult), parameter v1 and v2 become lower.

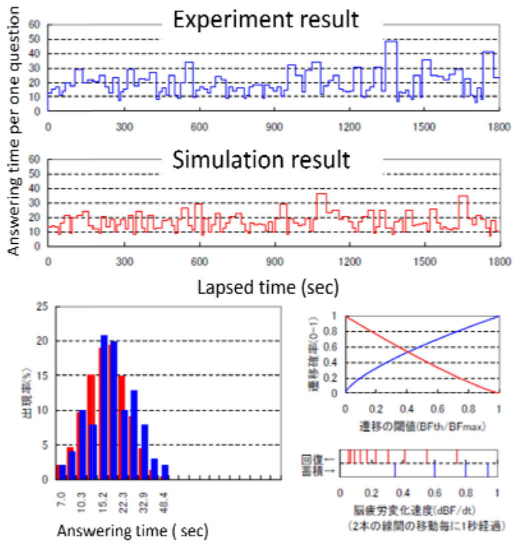


Fig. 7. Example of computer simulation result based on the model

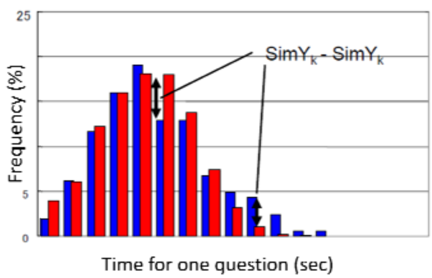


Fig. 8. Comparing the histogram of the simulation with experimental result

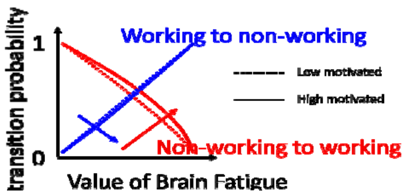


Fig. 9. Changes of state transition probability with motivation

Table 1. Model parameters and significance test for these parameters

	a_1			a_2			v_1			v_2		
Mental Work Load	Motivation			Motivation			Motivation			Motivation		
	Low	High	Variation	Low	High	Variation	Low	High	Variation	Low	High	Variation
Low MWL	0.93	1.03	0.10^{*}	0.93	0.81	-0.13^{**}	1.28	1.42	0.14	1.22	1.37	0.14
High MWL	0.99	1.07	0.08^{**}	0.85	0.64	-0.21^{**}	0.93	0.89	−0.03	0.92	0.89	−0.03
Both MWL	0.99	1.05	0.09^{***}	0.89	0.73	-0.17^{**}	1.11	1.16	0.06	1.07	1.13	0.06

	a_1			a_2			v_1			v_2		
Motivation	Mental Work Load			Mental Work Load			Mental Work Load			Mental Work Load		
	Low	High	Variation	Low	High	Variation	Low	High	Variation	Low	High	Variation
Low Motivation	1.03	0.85	-0.18^{*}	0.89	0.94	0.05	1.03	0.94	−0.09	1.01	0.9	−0.11
High Motivation	1.08	1	−0.08	0.83	0.72	−0.12	1.25	0.79	-0.45^{***}	1.2	0.78	-0.42^{***}
Both Motivation	1.06	0.93	-0.13^{*}	0.86	0.83	−0.04	1.14	0.87	-0.27^{***}	1.11	0.84	-0.27^{***}

*: $p<0.05$; **: $p<0.01$; ***: $p<0.001$

3.3.3 Discussion

By comparing the model parameter a_1 and a_2, it was found that the subjects tend to concentrate the task when their motivation is high. By comparing the model parameter v_1 and v_2, it looks the change speed of fatigue under high mental workload task is slower than low mental work load task. But high mental workload make the speed of accumulating fatigue slower looks opposite. Considering the change of working state simulated with the model, drop of the parameter v1 and v2 makes the time of a single working or non-working state longer. It means workers try to keep their attention to the task and take a long break instead when the task workload is high

4 Conclusion

In this study, the authors proposed a model which can explain and simulate the change of office work productivity. This model assumes office workers transit a working state and a non-working state with the probabilistic functions based on the variation of brain fatigue. Aiming to discuss the detail of the model, the authors have conducted a subject experiment in which work motivation are controlled. Comparing the results of the experimental results with the result of computer simulation based on the model, it was found that the subjects tend to concentrate the task when the remuneration is provided, and they try to keep their attention to the task and take a long break instead when the task workload is high.

Reference

1. Brill, M.: Using Office Design to Increase Productivity, vol. I. Buffalo Workspace Design and Productivity Inc. (1984)
2. Shimoda, H., Ito, K., Hattori, Y., Ishii, H., Yoshikawa, H., Obayashi, F., Terano, M.: Development of Productivity Evaluation Method to Improve Office Environment. In: 12th International Conference on Human-Computer Interaction, vol. 9(2), pp. 939–947 (2007)
3. Enomoto, K., Kondo, Y., Obayashi, F., Iwakawa, M., Ishii, H., Shimoda, H., Terano, M.: An Experimental Study on Improvement of Office Work Productivity by Circadian RhythmLight. In: The 12th World Multi-Conference on Systemics, Cybernetics and Informatics, WMSCI 2008, vol. VI, pp. 121–126 (2008)
4. Kondo, Y.: Development and Evaluation of Productivity Evaluation Method to Improve Office Environment. Magisterial Thesis, Kyoto University (2008)

Restoration Support System for a Historic Textile Market Using Virtual Environment

Michiko Ohkura , Mizuki Konuma, Yuri Kogure, Sayaka Tanaka, Hitomi Ei, Akiko Sakai, Takashi Ishidou, and Yoko Watanabe

Shibaura Institute of Technology,
3-7-5, Toyosu, Koto-ku, 135-8548 Tokyo, Japan
ohkura@sic.shibaura-it.ac.jp

Abstract. Recently, a trend is growing to restore such traditional Japanese buildings as theaters and markets as symbols of town renovation. However, the restoration of such buildings commonly encounters difficulties, including a lack of drawings of design and different construction materials and structures between the past and the present. Thus, virtual images of a restored building are a great help in evaluating and discussing restoration plans. Therefore, we constructed a restoration support system that uses virtual environments to help with the design plan for the indoor restoration of a historic textile market in Saitama Prefecture.

Keywords: traditional building, textile market, restoration, virtual environment.

1 Introduction

Recently, a trend is growing to restore such traditional Japanese buildings as theaters and markets as symbols of town renovation. However, the restoration of such buildings commonly encounters difficulties, including a lack of drawings of design and different construction materials and structures between the past and the present. Thus, virtual images of a restored building are a great help in evaluating and discussing restoration plans.

Based on this idea, we constructed a theater restoration support system that uses virtual environments to help the design plan for the indoor restoration of an old theater in Kawagoe City in Saitama Prefecture [1]-[3]. We constructed another restoration support system for the indoor restoration of an old storage facility and performed an evaluation experiment for it [4]. However, during the experiments, the following problem surfaced generational differences. They might surface during the selection of materials and colors for the interior of historical buildings.

We are constructing a new restoration support system using virtual environments for the design plan for the indoor restoration of the Old Kawagoe Textile Market. Fig. 1 shows a photograph of the original market from about 50 years ago.

We constructed a restoration system and performed some experiments. This paper describes the construction of the system and the experiments.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 413–422, 2011.

Fig. 1. Old Kawagoe Textile Market

2 Experimental Setup

The system diagram is shown in Fig. 2. We employed a 3D projector and LCD shutter glasses to stereoscopically show the interior of the virtual textile market.

An example of a room of the virtual textile market is shown on Fig. 3. Aged-brown wood was employed for the posts and beams. Wood and carpet were prepared as candidate materials for the floor. For the walls, mud and plaster were the candidate materials. The candidate colors for each material were based on the materials used in historical Japanese buildings.

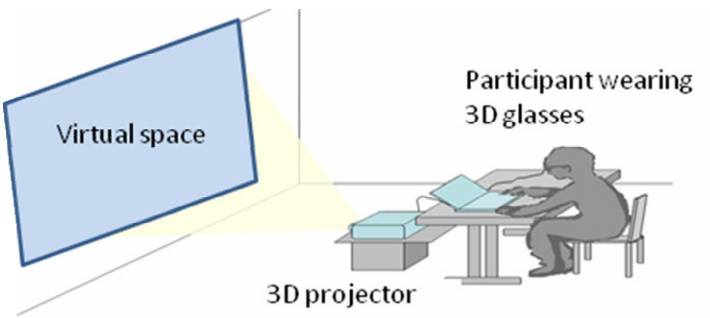


Fig. 2. Experimental setup

3 Preliminary Experiment

We performed a preliminary experiment for the following two reasons [5]:

- Because the number of candidate colors for each material was so many, we should reduce the number based on the results of the preliminary experiment.
- Because the age-dependency of the selection of colors was suggested in the experiment for the indoor restoration of an old storage facility [4], we should confirm that in case of the indoor restoration of the Old Kawagoe Textile Market.

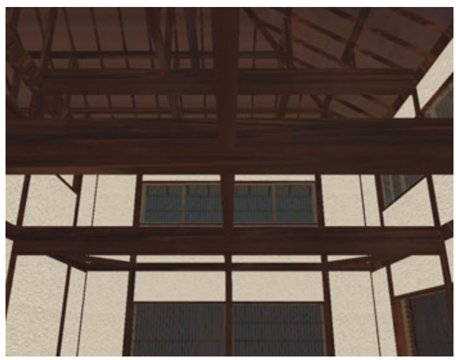
3.1 Experimental Method

Participants were presented a room of the virtual textile market with one of the combination of materials shown in Table 1, and asked to select the color of each material from the candidates shown in the menu windows. Fig.3 (a) is an example in which five candidate colors are presented for the floor's wood. The criterion of the color selection is to make the place most spacious. Fig.3 (b) shows another view of the virtual room used to confirm the selection. After selection, questionnaires gathered participant opinions about the important points of their selection of colors. This procedure was repeated four times for the four patterns in Table 1. The orders of presentation of the four patterns were counter-balanced. Finally, participants selected the most spacious combinations from the four patterns and explained their selection.

Table 1. Candidates of combinations of materials

	Pattern 1	Pattern 2	Pattern 3	Pattern 4
A) Floor	Wood	Wood	Carpet	Carpet
B) Walls	Mud	Plaster	Mud	Plaster

(a) front view



(b) upward view

Fig. 3. Examples of a room of virtual textile market

3.2 Experimental Results and Discussion

Experiments were conducted with 24 participants in their twenties, the young generation, and 24 participants in their forties or fifties, the middle-aged generation.

Figures 4, 5, 6 and 7 show the experimental results. Fig. 4 shows the frequency distribution of the selected colors of wood for the floor. Not all the color candidates for each material of the floor and walls were selected to emphasize spaciousness.

Fig. 8 compares the results between the selected colors for the wooden floor by generations. Table 2 summarizes the results of the most selected colors by generations for each material. From this table, we conclude that relatively bright colors were selected, except for the aged-brown wooden floors. In addition, the middle-aged generation tended to select brighter colors than the younger generation for all materials.

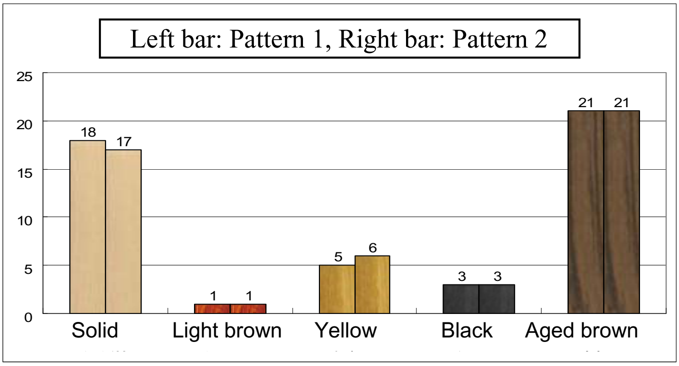


Fig. 4. Selected colors of wooden floor

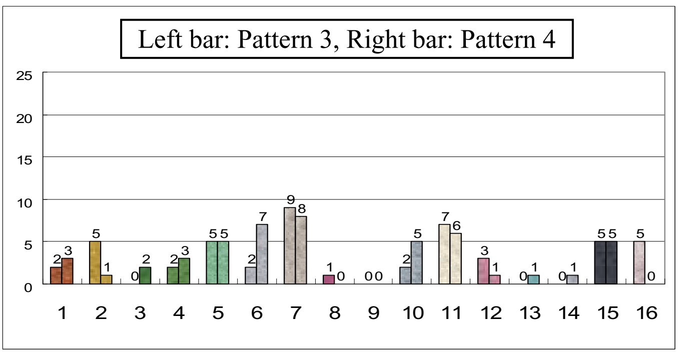


Fig. 5. Selected colors of carpet

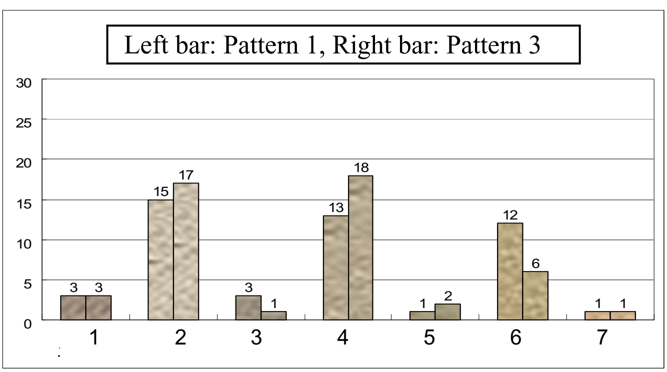


Fig. 6. Selected colors of mud for walls

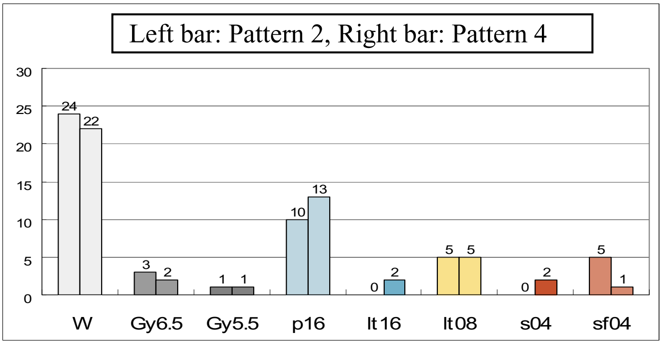


Fig. 7. Selected colors of plastered walls

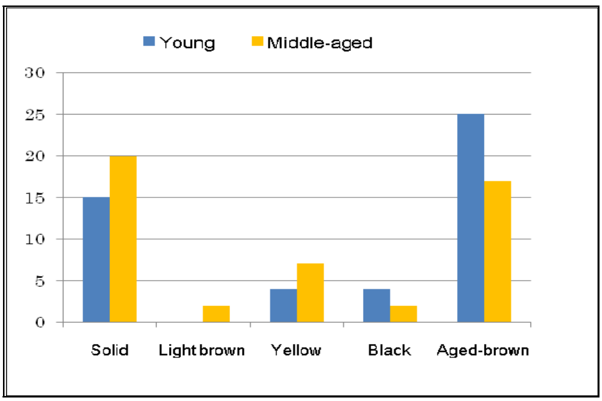


Fig. 8. Comparison between selected colors for wooden floor by generations

Table 2. Candidates of combinations of materials

Generation	Wood	Carpet	Mud	Plaster
Young	Aged brown	#7	#4	White
Middle-aged	Yellow	#11	#2	White

Figure 9 shows the result of the most spacious pattern. Pattern 2 was chosen most often by the young generation and second-most often by the middle-aged generation. Pattern 4 was chosen most often by the middle-aged generation and second-most often by the young generation. Since both Patterns 2 and 4 employed plaster as the material for their walls and white was selected most often both by the young and middle-aged generations, we conclude that white plaster is the most appropriate for the walls of this part of the textile market to create a spacious atmosphere.

Based on above results, the candidate colors for the next experiment were decided as shown in Table 3. In addition, we confirmed an age-dependency for the selection to emphasize the spaciousness of a virtual room of the old textile market.

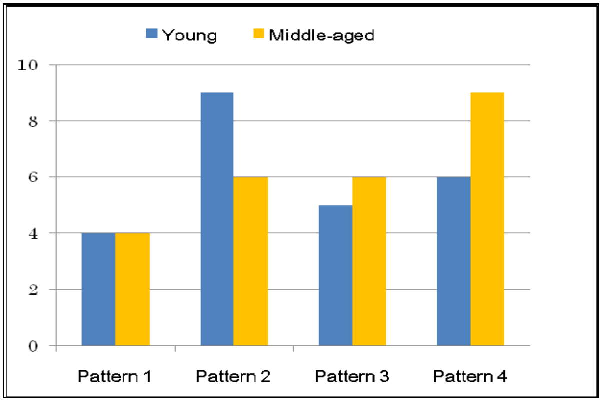


Fig. 9. Most spacious pattern

Table 3. Candidate colors for the next experiment

Material	Candidate Colors
Wood	Solid, Light brown, Yellow, Black, Aged brown
Carpet	Red, Yellow, Light green, Light gray, Beige, Light blue, White, Black, Purple
Mud	Brown, White, Beige, Yellow
Plaster	White, Light gray, Light blue, Yellow, Light red

4 Experiment

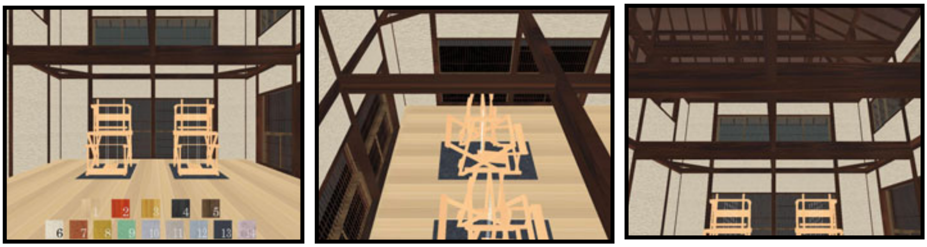
Employing the candidate colors shown in Table 3, we experimentally decided the materials and their colors for the walls and floors of a room of a virtual textile market. Because the room was planned to be used for a weaving experience, two weaving machines were placed on the carpet in the room (Fig. 10).

4.1 Experimental Method

The participants were presented a virtual room (Fig. 10), and asked to select the materials with colors from candidates in the menu windows. The selection was repeated three times for the wall, the floor, and the carpet under the weaving machines. The order of the selection part was set randomly. After the selection, questionnaires almost identical to the preliminary experiment gathered opinions about the important points of their selection. The selection criteria emphasized making the place spacious and nostalgic. The order of the criteria was counter-balanced.

4.2 Experimental Results and Discussion

The experiments were conducted at Kawagoe City with 32 participants in their twenties or thirties, the younger generation, and 32 participants from their forties to seventies, the older generation. In addition, the numbers of men and women were identical in both groups.



(a) front view (b) upward view (c) downward view from 2nd floor

Fig. 10. Examples of a room of virtual textile market

Figures 11, 12, 13 and 14 show the experimental results. Fig. 11 shows the frequency distribution of the selected materials with colors for the floor for spaciousness and Fig.12 shows that for nostalgia. In these figures, S, LB, Y, B, and AB represent Solid, Light brown, Yellow, Black, and Aged brown, respectively. Fig. 13 shows the frequency distribution of the selected materials with colors of the wall for spaciousness and Fig.14 shows that for nostalgia. Tables 4 and 5 summarize the results of the most selected combinations of materials with colors. From these results, we obtained the following:

- To emphasize spaciousness, relatively bright colors were selected both for the floor and the wall. The selected materials with colors were different between the generations; these results resemble those of the preliminary experiment described in 3.2.
- To emphasize nostalgia, relatively brilliant colors such as aged brown were selected both for the floor and the wall. The selected materials with colors were similar between the younger and older generations.
- No specific features were obtained for the selection of the carpet under the weaving machines.

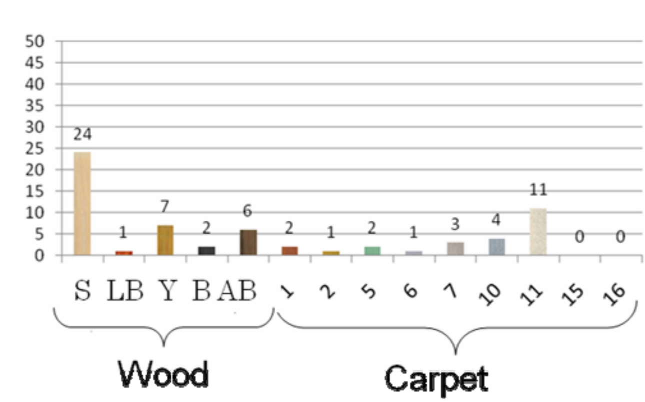


Fig. 11. Selected materials with colors of floor to emphasize spaciousness

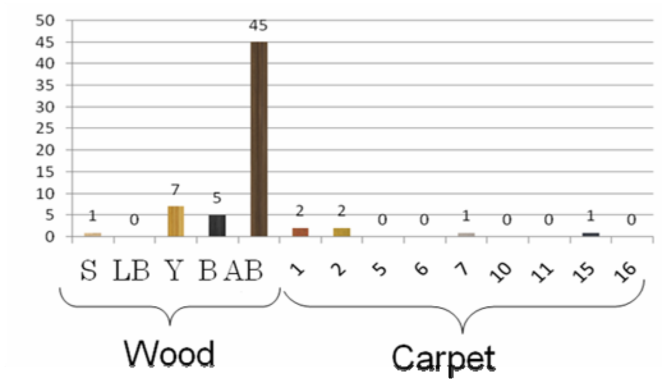


Fig. 12. Selected materials with colors of floor to emphasize nostalgia

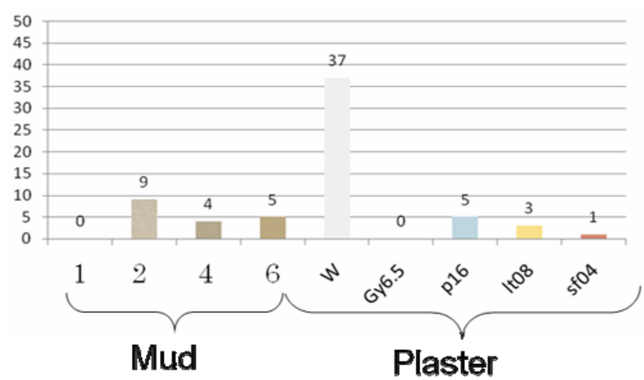


Fig. 13. Selected materials with colors of wall to emphasize spaciousness

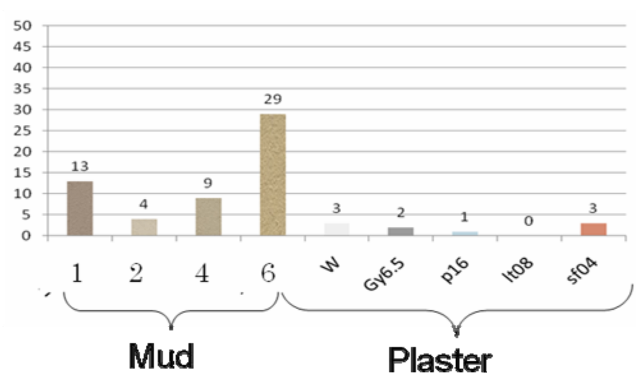


Fig. 14. Selected materials with colors of wall to emphasize nostalgia

Table 4. Experimental results for spaciousness

	Floor	Wall
Younger	Solid (25%), Carpet#11(19%), Aged brown (16%)	White (66%), Mud#2 (16%)
Older	Solid (50%), Carpet#11 (16%),	White (50%), Mud#6 (13%)

Table 5. Experimental results for nostalgia

	Floor	Wall
Younger	Aged brown (72%), Black (10%)	Mud#6 (53%)
Older	Aged brown (69%), Yellow (16%),	Mud#6 (38%), Mud#1 (28%)

These results, which show the differences and similarities of the selections for impressions emphasized by generations, suggest that the indoor restoration plan should be based on appropriate target impressions and target generation. The example of an indoor plan for spaciousness is shown in Fig. 15(a) and that for nostalgia is shown in Fig. 15(b).

5 Conclusion

We constructed a textile market restoration support system using virtual environments to help the design plan for the indoor restoration of an old Japanese textile market.

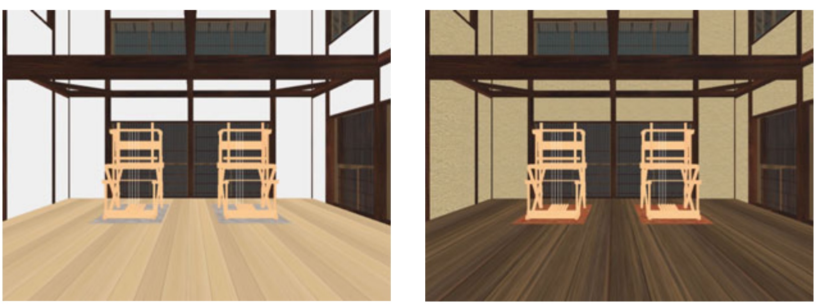


Fig. 14. Proposed indoor plan to emphasize spaciousness (left) and nostalgia (right)

Based on the results of preliminary experiments, the candidate colors for the next experiment were decided, and we confirmed an age-dependency for the selection to emphasize the spaciousness of a virtual room of an old textile market.

The results of experiments performed that selected materials and colors to emphasize impressions of spaciousness or nostalgia clarified that an appropriate indoor plan should vary and reflect the target impression and the target generations.

Acknowledgments. We would like to thank the visitors and the residents of Kawagoe City in Saitama Prefecture, and the Shibaura Institute of Technology students and staff who volunteered for the experiment.

References

1. Ohkura, M., Fujinuma, T., Mieno, H., Itoh, M., Watanabe, Y.: Theater Restoration Support System Using Virtual Environment. In: The 13th Intern. Conf. on Human-Computer Interaction, pp. 455–459 (2009)
2. Itoh, M., Mieno, H., Fujinuma, T., Ishidou, T., Ohkura, M., Watanabe, Y.: Method to Support Restoration of Historical Buildings Using Virtual Environment Study on Systematic Study on Theater Interior Through Impression Estimation. In: The 4th South East Asian Technical University Consortium (SEATUC) Symposium, pp. 20–23 (2010)
3. Itoh, M., Mieno, H., Fujinuma, T., Ohkura, M., Watanabe, Y.: Method to Support Restoration and Reconstruction of Historical Buildings using Virtual Environment - Study on Theater Interior through Impression Estimation. Transaction of Japan Society of Kansei Engineering 9(2), 161–170 (2010)
4. Itoh, M., Watanabe, Y., Ohkura, M.: Method to Support Restoration of Historical Buildings using Virtual Environment. Research report, 2009 Meeting of Kanto Branch of Architectural Institute of Japan (2010) (in Japanese)
5. Ohkura, M., Konuma, M., Kogure, Y., Ei, H., Sakai, A., Tanaka, S., Ishidou, T., Watanabe, Y.: Restoration Support System for a Historic Textile Market using Virtual Environment: Comparison between Generations, In: the 5th South East Asian Technical University Consortium (SEATUC) Symposium (to appear)

Part V

Design and Usability Case Studies

Investigating the Accessibility of Program Selection Menus of a Digital TV Interface

Pradipta Biswas and Patrick Langdon

Department of Engineering, Univiversity of Cambridge, UK
{pb400,pml24}@eng.cam.ac.uk

Abstract. We have investigated the accessibility issues of the program selection menus of a digital television interface. Initially we have simulated interaction patterns of visually and mobility impaired users and based on the simulation we have made changes to the font size and layout of the existing interface. Finally we have evaluated the new interface through a user trial involving people with disabilities. The results show that the new interface reduced the average time to select a channel and number of wrong channel selections for most of the participants.

1 Introduction

Many elderly persons and disabled people often get isolated from the society due to their disabilities and social issues. Television (TV) is a significant companion in their life to engage more fully with the world. Though there are legal, ethical and social reasons for designing products and services for people with different range of abilities, but still many people find the modern digital TV interfaces hard to use. We are working on the GUIDE project [20] that aims to develop a toolbox for providing adaptable, multi-modal TV interfaces and also helping designers to develop accessible interfaces for elderly and disabled users. In this paper, we have reported a study on accessibility issues of program selection menus of TV interfaces. We have identified the accessibility problems using a simulator and also used the simulator to propose new alternative design. Finally we have validated the simulation and new interface through a user study.

The paper is organized as follows. In the next section, we present background on the GUIDE project and our simulator. Section 3 presents related work followed by our study at Section 4. Finally we have drawn conclusions at Section 5.

2 Background

2.1 The GUIDE Project

The GUIDE project [20] is a medium scale focused research project on an EU FP7, Accessible and Assistive Information and Communication Technologies grant: “GUIDE – Gentle User Interfaces for Disabled and Elderly Citizens". It aims to develop a toolbox of adaptive, multi-modal user interfaces that target the accessibility

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 425–434, 2011.

requirements of elderly and impaired users in their home environment, making use of TV set-top boxes as processing and connectivity platform. For this purpose, the toolbox not only provides the technology of advanced multi-modal user interface components, but also the adaptation mechanisms necessary to let the UI components interoperate with legacy and novel applications, including the capability to self-adapt to user needs. The adaptation will be provided by modeling a wide range of capabilities of users. Figure 1 shows a schematic diagram of the GUIDE system. The GUIDE user model communicates with a middleware (the Adaptation Layer) of the GUIDE Hub which modifies application interfaces. The GUIDE user model uses a simulator to provide static (before interaction) and dynamic (during interaction) adaptation. The use of the user models and user profiles helps the system to address more varieties of users than existing similar system [9]. We shall use the following simulator for modeling users' interaction.

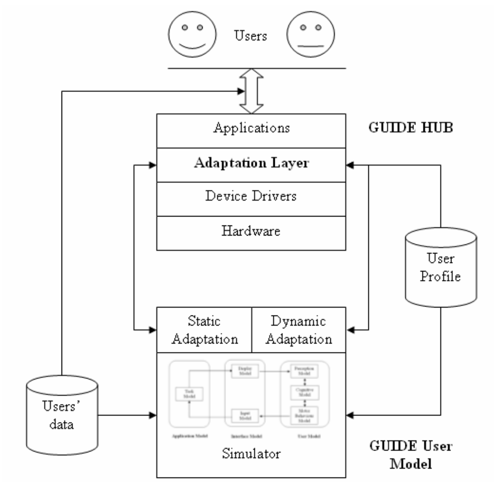


Fig. 1. Schematic overview of the GUIDE system

2.2 The Simulator

The simulator [4-7] embodies both the internal state of a computer application and also the perceptual, cognitive and motor processes of its user. It takes a task definition and locations of different objects in an interface as input. It then predicts possible eye movements and cursor paths on the screen and uses these to predict task completion times. The models are parameterized to represent different physical abilities, levels of skill and input devices. The simulator consists of a perception model, cognitive model and a motor behaviour model.

The perception model [5, 6] simulates the phenomena of visual perception such as focusing and shifting attention. It can also simulate the effects of different visual impairments (like Maccular Degeneration, Colour Blindness, Diabetic Retinopathy etc.) on interaction. We have investigated eye gaze patterns of able-bodied users as well as people with visual impairment and the model can predict the visual search time and eye gaze patterns of able-bodied people and a few types of visually impaired users (Myopia, Coour blindness and retinopathy [6]) with statistically significant accuracy. The cognitive model [4] simulates expert performance by using CPM-GOMS model [12]. It can also simulate performance of novices by using a dual space model [16]. The motor behaviour model [7] is developed by statistical analysis of cursor traces from motor impaired users. It quantifies the extent of impairment of a user by measuring his grip strength and uses it to predict movement time of a pointing task through a regression model. These models [4-7] do not need detailed knowledge of psychology or programming to operate. They have graphical user interfaces to provide input parameters and showing output of simulation.

3 Related Works and Our Approach

The GUIDE project plans to divide all possible sets of interaction with digital television into a finite set of scenarios and then investigating accessibility issues and providing adaptation for each scenario. This particular study considers a scenario of selecting channel and program from a menu. Previous work on menu selection investigated selection time of different menu items based on their position [15] and menu searching strategies [10] for able bodied users. Researchers worked on menu interaction for cell phones [13, 18] but there is not much reported work on accessibility issues of menus, in particular for digital TV interfaces. Existing approaches like target expansion [14] or target identification [11] are not very suitable for menu selection as menu items are more densely spaced than other types of targets like buttons or icons in a screen. Ruiz's approach [17] of expanding target region has also not been found to reduce menu selection time significantly. There is also not much reported work on the legibility issues of menu captions. Most researchers do not find difference in terms of reading time due to font types with respect to online reading tasks [1, 3, 8]. Though Bernard and colleagues [2] report significant difference in reading times between Tahoma and Corsiva fonts for a reading task of two pages, but the difference may turn insignificant during reading short captions. However, there is significant difference in reading time and legibility due to font size. Beymer and colleagues [3] prefer 12 point size while RNIB [19] and Bernard [1] prefer 14 point size.

Previous simulations [10, 18] on menu interaction mainly looked at cognitive aspects of interaction, but did not consider people with disabilities. We take help from our simulator [4-7] in identifying the accessibility problems of program selection menu with respect to visually and mobility impaired users. Based on the results of the simulation we have designed new interfaces. Our study consists of the following three stages

1. Problem identification through simulation
2. New interface evaluation through simulation
3. Validation of the simulation through a controlled experiment

4 The Study

Initially we have designed the following interface (Figure 2) which looks similar to existing systems (Figure 3). The GUIDE project [20] explores accessibility issues of people with a wide rage of abilities (including visual, cognitive and motor impairment) using different modalities of interaction (like pointing, keypad, gesture, voice based inputs and so on). In this particular work, we have investigated problems faced by people with visual and mobility impairment. Loss of visual acuity is one of the main symptoms of many visual impairments like Myopia, Maccular Degeneration, Diabetic Retinopathy and so on. Another significant type of visual impairment that affects vision without reducing visual acuity is colour blindness. Regarding mobility impairment, several types of motor impairment like stroke, cerebral palsy, dystonia or polio result significant spasm or tremor in finger which impedes users to use a remote control. However some of them can use a mouse, touch pad or trackball to move a pointer across the screen. So we have analyzed

- Sensory problems of
  - People with less visual acuity
  - People having colour blindness
- Interaction problems of
  - People with motor impairment using a pointing device

In this particular study, the simulator takes a sample task of selecting a menu item and the screenshot of the interface as input and shows the perception of visually impaired users and cursor trajectory of motor impaired users as output. In the simulation study, we have not bothered about the particular words used as captions since the simulation results are not to be used by participants. We use captions like Channel 1, Program 1 or Time 1 as captions. However in the validation study we used different words as captions and discussed it in detail in section 4.6.

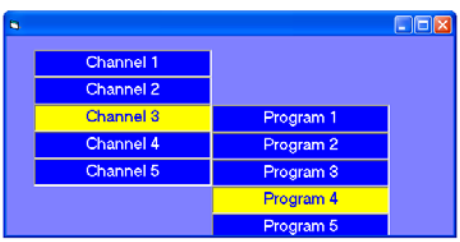


Fig.2. Control Interface

Fig. 3. Actual interface

4.1 Problem Identification

Initially the output from the simulator is used to identify accessibility problems. Based on the simulation results we identified the following two accessibility issues

- Legibility of captions
- Spacing between menu items

Figure 4 shows the redesigned interface. We have increased the font size of captions for users with visual impairment. For people with motor impairment, we have changed the size of the buttons without changing the screen size such as no couple of buttons shares a common boundary. It should reduce chances of missed clicks. We have also investigated the effect of severe visual acuity loss for the following six font types, however the legibility is not much different for different font types and nearly same for all.

Sans Serif	**Serif**
Microsoft Sans Serif	Sabon
Veradana	Times New Roman
Arial	Georgia

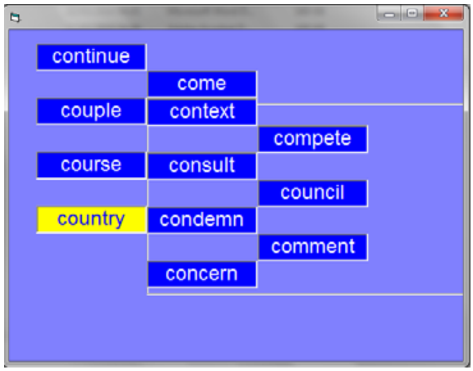


Fig. 4. New Interface

We evaluated the new interface with the following controlled experiment. In this experiment we hypothetize that

- People with visual acuity loss and motor impairment will perform a task faster and with less number of errors in the new interface (Figure 4) than the unchanged version (Figure 2).
- People with colour blindness will perform a task equally well with respect to people with no impairment (control group) in the unchanged version of the interface (Figure 2).

We measured the task completion time as a measure of performance and the number of missed clicks as a measure of errors.

4.2 Procedure

The procedure mimics the process of selecting a channel from a list followed by selecting a program from a drop down menu. Initially, the participants were shown a channel name and a program name. Then they made two selections matching the previously shown channel and program names. We did not use real channel and program names to avoid any biasness of users. The first two letters and length of all the captions were kept nearly same to avoid any pop-out effect [21] of the captions during visual search. We used the Veradana font type due to its bigger *x*-height and character spacing than other conventional fonts. Each participant repeated the task ten times. All participants were trained before undertaking the study.

4.3 Material

We used a standard optical Mouse and an Acer Aspire 1640 Laptop with a 15.5” monitor having 1280×800 pixel resolution. We also used the same seating arrangement (same table height and distance from table) for all participants.

4.4 Participants

We collected data from two institutes, National institute of Orthopedically Handicapped at Kolkata, India and Papworth Trust at Cambridge, UK. All participants have some experience of using computers- either they were learning or using computers regularly. All of them volunteered for the study.

4.5 Results

The average reaction time (total time needed to select the channel and program) was less in the new design than the control design (Figure 5) though the difference was not statistically significant in an independent sample two-tailed *t*-test ($t\ (120,1) = 0.64$, $p > 0.05$). The average number of missed clicks were also less (Figure 6) in the new design than the control design though the difference tends to statistical significance in a Wilcoxon ranked test ($W\ (120,1) = 163$, $p = 0.1$). In the experimental condition (new deign) missed clicks occurred in 21 trials while it occurred 31 times in control condition.

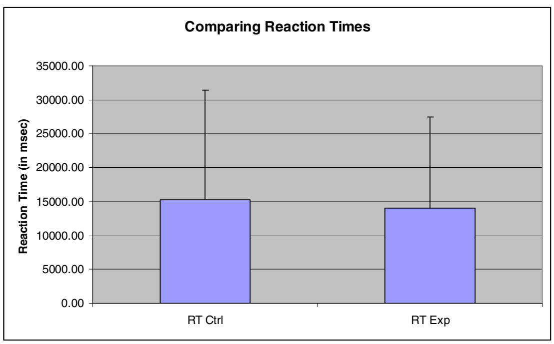


Fig. 5. Comparing reaction times

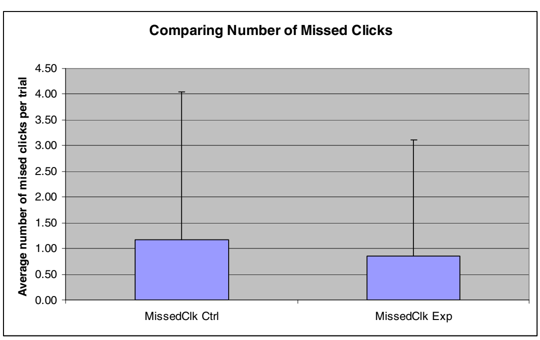


Fig. 6. Comparing number of missed clicks

We have also analyzed the reaction times and missed clicks for each individual participant. Table 2 shows the average reaction time and total number of missed clicks for each participant. It can be seen that 4 out of 12 participants (P4, P5, P8 and P9) have an average reaction time greater for the experimental condition and only 2 out of 12 participants (P8 and P12) missed clicked more in the experimental condition than the control condition.

4.6 Discussion

The reaction time and number of missed clicks were both less in the new design though we failed to find any statistical significance of the difference. Most of our participants do not have any problem in moving hands and thus they can control the mouse movement pretty well. Except participants P1, the visual acuity loss was also

Table 1. Participants

Participants	Age	Sex	Impairment
P1	>45	M	No mobility impairment. Age related Hypermetropia (+3.75 / +3.25 Dioptre)
P2	25-45	M	Difficulty in walking, right leg is shorter than left leg. Mild Myopia (- 2.75 / -2 Dioptre)
P3	25-45	M	Right hand was cut in accident, no impairment in left hand. No visual impairment.
P4	25-45	M	No mobility impairment. Lost vision in right eye, left eye is perfect.
P5	25-45	M	Left arm is affected by polio, no impairment in right hand. No visual impairment.
P6	<25	F	Lower body is affected by polio from birth, no impairment in hands, wheelchair user. No visual impairment.
P7	<25	M	Difficulty in walking from birth. Slight Myopia (-0.7 / -0.7 Dioptre)
P8	44	M	Cerebral Palsy reduced manual dexterity also some tremor in hand wheel chair user. Slight loss of visual acuity.
P9	63	M	Left side (non dominant) paralysed after a stroke in 1973 also has tremor
P10	31	M	Cerebral Palsy reduced manual dexterity wheel chair user.
P11	>45	M	Reduced manual dexterity in limbs due to neurological problem, wheel chair user.
P12	44	F	Did not mention disease restricted hand movement no tremor. Slight loss of visual acuity.

Table 2. Result per participant

	AvgRT Ctrl	**AvgRT Exp**	**TotalMC Ctrl**	**TotalMC Exp**
	(in msec)			
P1	3886	3259	0	0
P2	5755	5033	0	0
P3	7230	6149	0	0
P4	21777	26838	72	56
P5	4481	4611	0	0
P6	12195	11739	11	4
P7	15628	6747	13	0
P8	15394	18628	20	28
P9	7213	9184	0	0
P10	36160	25084	11	0
P11	20752	20550	14	8
P12	32228	30223	0	6
Avg	**15225**	**14004**	**11.8**	**8.5**

not severe. Additionally in the present experimental set up, a missed click did not waste time while in a real interface a missed click will take the user to an undesired channel and getting back to the previous screen will incur additional time. So the higher number of missed clicks in the control condition will also increase the channel selection time further in an actual scenario. However in future we plan to run the study with more cautious selection of participants. All of the visually impaired participants preferred the bigger font size. However a few participants reported difficulty in reading the zigzag presentations of captions of the new interface. In future we also plan to use an eye tracker to compare the visual search time for both types (linear and zigzag) of organizations of menu captions.

In this study, we have investigated the accessibility of program selection menus for a digital TV interface. Future studies will include more interaction modalities (like keypad or gesture based interaction), devices (like remote control, set top box and so on) and impairments (like cognitive impairments). However the results of this study can be extended beyond program menu interfaces of digital televisions. For example, the font size of captions in absolute terms (x-height ≈ 0.5 cm.) indicates the minimum font size required for any text in an interface for serving people with severe visual acuity loss. Similarly the particular colour combination of the screen (white text in blue background) can be used in any other interface as well to cater people with colour blindness. Finally the modified menu structure can be used in computers or other digital devices to make the menus accessible to people with mobility impairment.

Unfortunately we did not get any participant with colour blindness. So we have used a colour blindness filter (from Cambridge Research Systems, http://www.crsltd.com) to simulate the effect of dichromatic colour blindness. In this case, we did not find any significant difference in reaction times in an independent sample two-tailed t-test (t (20,1) = 0.81, p > 0.05)and did not record any missed clicks as well.

5 Conclusions

In this paper we have investigated the accessibility issues of digital television interfaces. We simulated interactions of visually and mobility impaired users. Based on the simulation, we recommended change of the layout and font size of menu captions and designed a new interface. Finally, we evaluated the new interface through a user trial involving users with visual and mobility impairment. Though we failed to find any statistically significant difference between the existing and new design but the new interface reduced the average time to select a channel and number of wrong channel selections for most of the participants.

References

1. Bernard, M., Liao, C.H., Mills, M.: The Effects of Font Type and Size on the Legibility and Reading Time of Online Text by Older Adults. In: ACM/SIGCHI Conference on Human Factors in Computing Systems 2001, pp. 175–176 (2001)
2. Bernard, M., Mills, M., Peterson, M., Storrer, K.: A Comparison of Popular Online Fonts: Which is Best and When? Usability News (2001)
3. Beymer, D., Russell, D.M., Orton, P.Z.: An Eye Tracking Study of How Font Size, Font Type, and Pictures Influence Online Reading. In: 22nd British Computer Society Conference on Human-Computer Interaction, pp. 15–18 (2007)
4. Biswas, P., Robinson, P.: Automatic Evaluation of Assistive Interfaces. In: ACM International Conference on Intelligent User Interfaces (IUI) 2008, pp. 247–256 (2008)
5. Biswas, P., Robinson, P.: Modelling Perception using Image Processing Algorithms. In: 23rd British Computer Society Conference on Human-Computer Interaction, HCI 2009 (2009)
6. Biswas, P., Robinson, P.: Modelling user interfaces for special needs. In: Accessible Design in the Digital World, ADDW (2008)
7. Biswas, P., Robinson, P.: Predicting pointing time from hand strength. In: Holzinger, A., Miesenberger, K. (eds.) USAB 2009. LNCS, vol. 5889, pp. 428–447. Springer, Heidelberg (2009)
8. Boyarski, D., Neuwirth, C., Forlizzi, J., Regli, S.H.: A Study of Fonts Designed for Screen Display. In: ACM/SIGCHI Conference on Human Factors in Computing Systems, pp. 87–94 (1998)
9. Epelde, G., et al.: URC based Accessible TV. In: EuroITV 2009, pp. 111–114 (2009)
10. Hornof, A.J., Kieras, D.E.: Cognitive Modeling Reveals Menu Search is both Random And Systematic. In: ACM/SIGCHI Conference on Human Factors in Computing Systems, pp. 107–114 (1997)

11. Hurst, A., Hudson, S.E., Mankoff, J.: Automatically identifying targets users interact with during real world tasks. In: ACM International Conference on Intelligent User Interfaces (2010)
12. John, B.E., Kieras, D.: The GOMS Family of User Interface Analysis Techniques: Comparison And Contrast. ACM Transactions On Computer Human Interaction 3, 320–351 (1996)
13. Marsden, G., Jones, M.: Ubiquitous Computing and Cellular Handset Interfaces – are menus the best way forward? In: South African Institute of Computer Scientists and Information Technologists Annual Conference (2001)
14. McGuffin, M., Balakrishnan, R.: Fitts' law and expanding targets: Experimental studies and designs for user interfaces. ACM Transactions on Computer-Human Interaction 12(4), 388–422 (2005)
15. Nilsen, E.L.: Perceptual-motor Control in Human-Computer Interaction. The Cognitive Science and Machine Intelligence Laboratory, the University of Michigan, Ann Arbor, MI (1992)
16. Rieman, J., Young, R.M.: A dual-space model of iteratively deepening exploratory learning. International Journal of Human-Computer Studies 44, 743–775 (1996)
17. Ruiz, J., Lank, E.: Speeding Pointing in Tiled Widgets: Understanding the Effects of Target Expansion and Misprediction. In: ACM International Conference on Intelligent User Interfaces, pp. 229–238 (2010)
18. St. Amant, R., Horton, T.E., Ritter, F.E.: Model-based Evaluation of Expert Cell Phone Menu Interaction. ACM Transactions on Human-Computer Interaction 14(1), 1–24 (2007)
19. See it Right. UK: RNIB (2006)
20. The GUIDE Project, `http://www.guide-project.eu/` (accessed on August 23, 2010)
21. Treisman, A., Gelade, G.: A Feature Integration Theory of Attention. Cognitive Psychology 12, 97–136 (1980)

Windows Positioning System: Aural Assistance Environment for the Aging in Windows Navigation

Chi Nung 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 study explored dual-sensory user interface designs that meet the requirements of elderly people. Eight subjects with conventional mouse and the Windows Positioning System interface processing strategies were conducted the experimental tasks where five types of applications were evaluated in different font sizes. The analysis illustrates that for the conventional mouse the difference of response time of each font size of target texts is significant by ANOVA to the five font sizes 8, 10, 12, 14, and 16. However, the difference between large font sizes of 14 and 16 was not significant except the result between font sizes of 10 and 16. But for the WPS, there is no difference of response time among the five font sizes. It showed that the effect of aural information of target texts can benefit user from computer manipulation.

Keywords: Windows Positioning System**,** Text-to-Speech Engine.

1 Introduction

Access to the Internet has emerged as an integral part of human society. For the aging society coming, however, there are some requirements of computer windows manipulation faced with the users in low vision. The elderly people are often less comfortable and perceive themselves as having less control over computers for typical, computer-related tasks [1, 2, 3]. Since a majority of computer-use time is occupied through the use of reading, the presentation of text is important for the elderly people. Although the elderly people perceived 14-point fonts as being more interesting than 12-points font, the various factors associated with age [4]. For assisting the aging with low vision in computer windows navigation, other than visualization hearing sensory information has become another important orientation in computer windows environment. Because it can tell where one stays, where one is going, and read out to identify what the program or object is pointed by the cursor.

2 Windows Positioning System

The design of Windows Positioning System (WPS) (Fig. 1) provides an aural assistance environment which integrates the IBM Chinese Text-to-Speech Engine and

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 435–438, 2011.

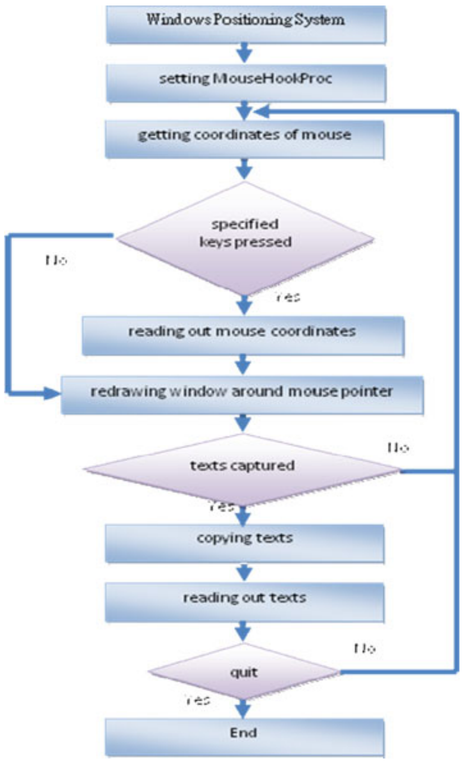


Fig. 1. Windows Positioning System mechanism

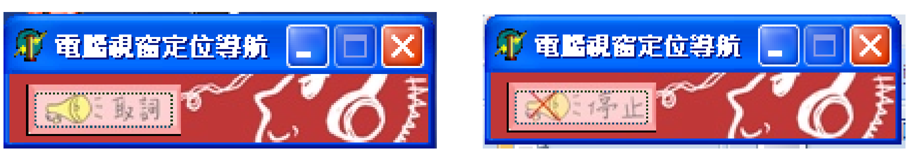


Fig. 2. Switching Talking Aid function of Windows Positioning System

technology of capturing screen texts. The WPS focuses on two functions: Talking Aid (Fig. 2) and Cursor Positioning. On the Talking Aid aspect, WPS once triggered is able to read out the texts nearby where the cursor was just moved to. This would confirm users by the voice of IBM Chinese Text-to-Speech Engine stating the associated texts of an application function, a file folder, a computer object, or contents within an application container. Therefore, WPS could identify users with aural information of any object target in computer windows navigation. At the same time users could also highlight the texts of contents in the opened application or a file for reading out by the WPS. On the Cursor Positioning aspect, WPS can bring users to specifying cursor position by the readout of cursor axes in x-coordinate and y-coordinate, aiming to facilitate the positioning of mouse. Thus while referring to the Cursor Positioning function of WPS, users can distinguish where the mouse cursor is in the computer windows environment.

3 Evaluation of Windows Positioning System

Subjects included in the study were eight elderly users with average age 68 years old. They all have the experience of computer manipulation for over 3 years for daily usage. In the experiment, subjects were asked to use the conventional mouse and the WPS separately to launch everyone's own routines task and new unfamiliar task on the Microsoft Windows platform. The WPS which was developed by Delphi 6.0 is able to monitor the mouse cursor moving and provide corresponding aural information triggered by the function of Talking Aid or Cursor Positioning. The target task texts appear 5 different kinds of font sizes, 8, 10, 12, 14, and 16. Experiments was conducted with a counterbalanced measures design to avoid the pitfalls of standard repeated measures designs, where the subjects were asked to launch one's own familiar task and unfamiliar task interlacedly with conventional mouse and the WPS separately.

Experiment compared the time between the two kinds of familiar and unfamiliar tasks at movement time, the font size variation of target texts by ANOVA and regression analysis. Results show that response time of unfamiliar tasks tends to be longer than the familiar ones for both conventional mouse and WPS. And the response time tends to decrease as the use of WPS compared with use of conventional mouse. The results also show that the response time decreases as the font size of target become larger. The analysis illustrates that for the conventional mouse the difference of response time of each font size of target texts is significant by ANOVA to the five font sizes 8, 10, 12, 14, and 16. However, the difference between large font sizes of 14 and 16 was not significant except the result between font sizes of 10 and 16. But for the WPS, there is no difference of response time among the five font sizes, thus it can be said that the effect of aural information of target texts can benefit user from computer manipulation.

4 Conclusions

The WPS works with an aural assistance environment to allow users to select either the appropriate functions of Talking Aid and Cursor Positioning to navigate where they are in the computer windows. A GPS-like loudspeaker in the computer windows environment, WPS can help guide the aging with low vision by hearing sensory information to facilitate identifying where they are in the computer world.

References

1. Bernard, M.L., Liao, C.H., Mills, M.M.: The Effects of Font Type and Size on the Legibility and Reading Time of Online Text by Older Adults. In: Proceedings of CHI 2001 Extended Abstracts, pp. 175–176 (2001)
2. Czaja, S.J., Sharit, J.: Age Differences in Attitudes Toward Computers. Journal of Gerontology. Psychological Sciences 5(5), 32–340 (1998)
3. Dyck, J.L., Smither, J.A.: Age Differences in Computer Anxiety: The Role of Computer Experience, Gender and Education. Journal of Educational Computing Research 10(3), 239–248 (1994)
4. Vanderplas, J.M., Vanderplas, J.H.: Some factors affecting legibility of printed materials for older adults. Perceptual and Motor Skills 50, 923–932 (1980)

User Interactive Design for Digital TV Web Surfing

Chih-Fei Chuang, Nancy Huang, and Sheue-Ling Hwang*

Department of Industrial Engineering & Engineering Management,
National Tsing-Hua University, Hsin-Chu, 300, Taiwan
No. 101, Kuang Fu. Rd., Sec. 2, Hsinchu, 300, Taiwan
Tel.: +886-3-5742694; Fax: +886-3-5722685
slhwang@ie.nthu.edu.tw

Abstract. There is no standard for website layout and design of remote control device to date, and thus this study aims at proposing evaluation and suggestion to website layout as well as designing a proper and satisfactory remote control device for Web TV through human factors methodology, thereby to meet the demand of television history. There were two experiments in this study. Experiment I focused on website layout design for Web TV, investigating the proper sizes of characters and cursor as well as finding out possible gesture control motions. On the other hand, Experiment II focused on the design of remote control device, looking for the operating function for Web TV by means of motion analysis, usability evaluation and competitive advantage analysis for five different devices. After integrating analysis of all experimental data, we proposed a default value of the size of characters and cursor for the website layout design; besides, we provide suggestions for the design of remote control device and also design a new concept remote control device.

Keywords: Interactive design; Digital TV; Web surfing.

1 Introduction

Web TV is neither television nor computer; it is a new convergent medium. Therefore, it should have its own right and standards for the programming and interaction. Since the standard on the website layout and the design of remote control device have not been established to date, this research aims at proposing suggestions to website layout and designing a proper and satisfactory remote control device for Web TV through human factors methodology, thereby to meet the demand of television history.

Chorianopoulos and Lekakos[1] define interactive TV (iTV) as a user experience that involves at least one user and one or more audiovisual and networked devices. A Web TV is a television set especially designed (or connected using a set-top box) to allow an Internet connection [2].Current user interfaces for iTV(interactive television) consist of graphical representations of services projected on the TV screen and

* Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 439–442, 2011.

manipulated by means of the remote control [3]. The keypad of the remote control device consists of labeled buttons that are mapped to specific TV control actions, such as for channel up/down and mute. For example, to select a channel, the viewer can enter the number of the channel, use the channel up/down button, or navigate through the graphical representation. Viewers need either to be aware of the channel position, or to search for it. This is problematic, in particular when many channels are available, because remote controls are ill-designed for complex interaction [4].

Now several TV functionalities not only require easy selecting processes but also complex entries including pointing or text input. Therefore, Today's standard remotes are less qualified for universal multimedia operations [5].

2 Method

There were three tasks in Experiment I.

Task 1: The goal of task 1 was to find out the proper size of character for Web TV. All participants watched TV showing three main websites associated with music and video, e-mail, and climate information, and adjust the size of characters in the web through the function of Zoom in/out.

Task 2: The purpose of task 2 was to find out the proper size of cursor for Web TV. All subjects were asked to pick out the favorable size of cursor under different condition of website layouts as task 1.

Task 3: The aim of task 3 was to investigate the possible gesture control for Web TV. All participants needed to act it out supposing they were able to control the website via their gestures.

There were thirty participants in this experiment. Fifteen participants were between 22 and 28 years old. Another fifteen participants were between 38 and 66 years old.

Before Experiment II, we need to find out the essential function while operating the web TV. In the experiment II, we evaluate different remote control devices through methodology of human factors, thereby understanding competitive advantage of these different remote control devices. From the result of motion analysis of Web on TV, there are four essential operating functions including Click, Zoom in/out, Movement of cursor, and Movement of web page. Then, Experiment II carries out subjective and objective usability evaluation to estimate these different remote control devices including EeePC, Play Station 3, Wii, Pen Tablet, and Touchpad.

3 Results

3.1 Experiment I

Result of the proper size of cursor for Web TV is 2.32×1.39 cm (Length×width) , and the proper size of character depends on different conditions. There are two kinds of conditions: one is that when the website is full of text, the proper size of character is

1.43×1.85 cm; the other is when the website includes lots of pictures/figures, the proper size of character is 1.19×1.65 cm.

The result of t-test for the outcome between young and elder groups showed that there is no significant difference between these two groups in the proper size of cursor ($p=0.185$) and character ($p=0.279$).

3.2 Experiment II

For operating time, the result of ANOVA showed that there was no significant difference in operating time among different devices as a whole ($p=0.335$) and for the young ($p=0.751$). However, there was a significant difference among different devices for the elder. The operating times between EeePC and PS3 were significantly different. Therefore, EeePC was more efficient than PS3 for the elder.

For texting time, the result of ANOVA showed that there was significant difference in operating time among different devices. Texting times were significantly different among Wii, EeePC, TouchPad and PS3 as well as Pen Tablet for the young; and significantly different among EeePC and Wii as well as PS3 for the elder. Overall, texting times of EeePC and Wii were significantly faster than that of Pen Tablet and PS3; therefore EeePC and Wii were more efficient than Pen Tablet and PS3 in terms of texting.

4 Discussion and Conclusion

4.1 Default Value for Website Layout

From the result of t-test for the comparison between young and elder groups, we can understand that there is no significant difference between these two groups regarding to the proper size of cursor and character, hence we can propose a default value for the size of cursor and character on Web TV. The standard size of the cursor is 2.32×1.39 cm (Length×width), and the standard size of character is 1.19×1.65 cm when the website is including lots of pictures/figures and 1.43×1.85 cm while the website is full of text.

4.2 Suggestions for Remote Control Design

1. The remote control device should be simple and its operating functions should be easy and intuitive to user.
2. The posture while using the device should be taken into consideration in order to avoid body aches caused by uncomfortable posture.
3. The view sight should be considered while designing the remote control in order to minimize the frequency of sight switch.
4. The problem of accidental press should be considered while designing the remote control device.

5. We recommend using a single key for the function of Click. We do not suggest the Click function of EeePC because of the double screen would compel the subject to switch their sight. Additionally, using TouchPad and Pen Tablet might cause some problems such as accidental press.
6. We recommend using the individual zoom keys or the touch ring for the function of Zoon in/out. The way of using two fingers to zoom in/out is not recommended because of accidental press problem.
7. We recommend two ways for the function of Movement of cursor: one is just to hold the device in the air and wave; another is to move single finger on the touchpad.
8. We recommend using one finger sliding on the touchpad for the function of Movement of web page.

By summarized analysis and pro and con discussed above, we provide a new concept for a remote control device which is shown below.

The core advantages of the new remote control device are listed below:

1. It is one-hand remote control device.
2. The operating function is simple and intuitive, thereby easy to be understood.
3. The control device owns independent Click key; therefore there is no need to worry about the problem of accidental press.
4. There are two modes for Movement of cursor, so users do not have to hang on the same gesture persistently, and they can use different gesture to control the Web TV. Thus the device would not lead to body aches for prolonged use.
5. The device is a low costly touch screen. Without double screens, users do not have to lower and raise their head frequently which might causes neck aches.

Acknowledgments. This study is supported by Media Tek- NTHU joint project

References

1. Chorianopoulos, K., Lekakos, G.: Learn and play with interactive TV. ACM Computers in Entertainment 5(2), Article 4 (2007)
2. Lu, K.J.: Interaction Design Principles for Interactive Television. Unpublished Masters Thesis, Georgia Institute of technology (2005)
3. Eronen, L., Vuorimaa, P.: User interfaces for digital television: a navigator case study. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 276–279. ACM Press, Palermo (2000)
4. Gill, J.M., Perera, S.A.: Accessibility of universal design of interactive digital television. In: Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors (EuroITV), pp. 83–89 (2003)
5. Hess, J., Küstermann, G., Pipek, V.: pRemote - a user customizable remote control. In: ACM Conference on Human Factors in Computing (CHI 2008), pp. 3279–3284. ACM Press, Florence (2008)

An End User and Environment Field Study for an Inclusive Design of Consumer Products

Thomas Fiddian[1], Chris Bowden[1], Mark Magennis[1],
Antoinette Fennell[2], Joshue O' Connor[2], Pierre T. Kirisci[4], Yehya Mohamad[3],
and Michael Lawo[4]

[1] The Royal National Institute for Deaf People LBG (RNID), Featherstone Street 19-23, EC1Y 8SL London/UK
[2] National Council for the Blind (NCBI)
[3] Fraunhofer Institute for Applied Information Technology (FIT)
[4] Universität Bremen
Thomas.Fiddian@rnid.org.uk

Abstract. In this paper an approach to improve the design of every day consumer products for inclusive design with a focus on elderly people with mild to medium physical and sensory impairments is presented. As mainstream manufactures do not have a detailed understanding of the needs of this target group the idea is to use a Virtual Human Model that covers these impairments. A Virtual Laboratory with three design phases is the approach to allow designers to plan and evaluate the user interfaces of their products. The paper gives a state of the art and presents the Virtual User Model as a mixture of human and environment context. In this paper we present results of an detailed ethnographic study. The research carried out on a group of 58 elderly people from the UK, Ireland and Germany who had a range of three mild-to-medium impairments; hearing, vision and manual dexterity.

1 Introduction

If all products and services could be designed for inclusion, most barriers that physically impaired citizens confront in our society would be overcome. But how can consumer-oriented companies adapt their existing product portfolios to the needs of sensory and physically impaired users in a competitive and sustainable manner? How can product developers be empowered during product development to consider the real-time accessibility needs of physically and sensory impaired users, with the aim of securing universal access to future consumer products?

The needs of people with impairments are still not considered sufficiently well when designing user interfaces (UIs) for accessing and interacting with consumer products. Currently available user interfaces and interaction devices rarely fulfil the real-time interaction and accessibility requirements of users with vision, hearing, and manual dexterity impairments. Thus it is not uncommon for an individual to have multiple impairments, which is particularly common among elderly people. The great majority of people older than 70 years have mild to medium age-related impairments (hearing reduction, macular degeneration, etc) rather than profound impairments.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 443–452, 2011.

They can only fully benefit from consumer products, when UIs incorporate accessible multimodal interaction capabilities providing good usability, and thus possess some degree of context sensitivity. Due to the complexities of singular and multiple age-related impairments, it is unrealistic for a mainstream manufacturer to have a detailed understanding of these and design appropriately. Therefore only by support from a third party can form first party knowledge.

"User experience" is a growing industry factor with an increasing number of manufacturers indentifying it as a way to improve the quality of their product line. Infotainment devices such as digital cameras, mobile phones, TV and radio, as well as white goods such as household appliances (fridges, cookers, etc.) could benefit if designed in a way usable also by users with mild to medium age related impairments.

2 Objective and Significance

User interfaces of every day products – such as washing machines, ovens, dishwashers, telephones, televisions and so on – have become increasingly complex, moving away from simple mechanical buttons and controls towards detailed and complex digital displays [1].

Age affects our vision, hearing, manual dexterity, mobility, ability to remember or concentrate, strength, stamina and so on. This is likely to affect how we interact with products. Most products – except perhaps those that are specifically designed for older people – are developed without any consultation with older users during the design process. As a result, designs that require a certain level of technological expertise or that require unnecessarily complex interfaces have become the norm and many older people are being excluded from using them.

To identify and describe key usability issues that people encounter when using specific consumer product types we focused upon people with one of three common impairments and people showing combinations of these impairments. One group was of users with one minor developed physical impairment such as visual impairment, auditory impairment, and manual dexterity impairment. Another one was of elderly users with age related impairments – usually a combination of mild to medium of the previously mentioned impairments. These can also be referred to as multiple sensory impairments.

The levels of impairment severity covered by our research were mild to medium (as opposed to severe or profound) and these were determined for each participant during the research process. The mentioned impairments were chosen because of the commonality of the afflictions, the effect it has on using consumer products (touch, sight and hearing, are the primary senses used when interacting with an inanimate object) and the partners involved in the project (RNID are experts in hearing loss, NCBI are experts in visual loss). The consumer products to be investigated in detail were washing machines and mobile phones.

The research involved carrying out detailed ethnographical studies of the participants in their own home. The most important aspects of this research was identifying key problem areas, looking for commonality within and between impairment groups and differing products, and presenting this information in an accurate but accessible and usable format.

3 Method

Detailed ethnographic research was carried out on a group of 58 elderly people from the UK, Ireland and Germany who had a range of mild-to-medium impairments. Three types of WHO classified impairments [2] were focused on; hearing loss (B230), sight loss (B210) and manual dexterity (B710/730). The research comprised of a combination of interview and observational techniques and investigated the main usability problems which these specific users encountered when using their washing machine and mobile phone in a typical use environment.

4 Encountered Impairments

We focused on two groups of people, users with one minor developed physical impairment such as visual impairment, auditory impairment, and manual dexterity impairment and elderly users with age related impairments – usually a combination of mild to medium above mentioned impairments. These two groups of people are not mutually exclusive, due to the type and severity of the impairments that the field trial is focused on; minor or mild to medium visual, auditory and manual dexterity impairment. These three impairments are all far more common with older people (65+) and the most common variants of these impairments are directly related to ageing; such as presbyacusis which is age related sensorineural hearing loss. For these reasons this study will be biased towards age related types of visual, auditory and manual dexterity impairments, and elderly users. We reviewed various types of hearing, vision, and manual dexterity impairments.

The level of impairment we covered is defined as "mild to medium". Medium hearing impairment is not a recognized category of hearing loss, however as one of 'Moderate's dictionary definitions is "of medium quantity, extent, or amount "[1], the study defines a medium hearing loss as the same as moderate hearing loss (see fig. 1).

Hearing impairment Descriptors	Definitions of hearing loss	Hearing Loss (dB)*
Mild impairment	People who suffer from mild hearing impairment have some difficulty following speech and keeping up with conversations, especially in noisy surroundings.	20 – 40
Medium impairment	People who suffer from moderate hearing impairment have difficulty following speech keeping up with conversations when not using a hearing aid or other amplification equipment.	41 – 70

Fig. 1. Hearing loss definition

**Audiometric Descriptors are based on the average of the pure tone hearing threshold levels at 250, 500, 1000, 2000 and 4000 Hz; 0dB reference based on assessments of a large group of people considered as having 'normal' hearing*

---

[1] Moderate: "of medium quantity, extent, or amount: *a moderate income."*
http://dictionary.reference.com/browse/moderate

Any selected method of vision impairment assessment would be hampered by the variation in testing conditions, as the field trials took place in each of the user's homes. In order to be accurate, the standard Snellen test must be performed at a particular distance from the eye chart (which a user's house is not likely to cater for) and at a particular light level (which is impossible to control for at the user's home). As a result, a simple reading test was developed to allow the authors to categorise or distinguish between users, rather than to quantify level of vision loss.

Methods of assessing manual dexterity include focusing on gross and fine manual dexterity, motor coordination, grip strength, tactile recognition, two-point discrimination, and touch/pressure thresholds, among others.

Unlike hearing and vision, manual dexterity cannot be easily measured in the field in a quantitative manner. Due to the large degree of variation in the above methods, and the inability of these assessments to consider factors such as pain, discomfort, weakness, and environmental factors such as temperature, it was decided to perform a subjective, rather than objective, investigation of manual dexterity. Observations of the user's manual dexterity (where particular limitations of movement, or obvious physical impairments were observed) were also noted by the investigators.

5 Methodology

Our methodology for the impairment assessment and the field study was carried out in the context of an EC funded research project [3][2]. The main research methodology employed was detailed observational studies carried out in the participants own home environment. This methodology was used because when looking at the main usability problems associated with specific products, an inherently qualitative study, it is vitally important to reduce the number of experimental errors. As a result it is important to only use products that the users are familiar with and that they have already learnt to operate, and that they use such products for tasks that they regularly carry out and in environments that are familiar to them. The importance of using the participants own products are that they have already had sufficient time to learn how to use them, so any usability problems that now exist will be significant and ones which the user has not been able to overcome easily. If they were asked to operate and assess a totally new product then the main part of the research would focus upon how intuitive the product is, how quickly a user can learn to use the product and how good any instructions are. Similarly, if you introduce the user into a laboratory environment, or one which they are unfamiliar with, you will immediately be introducing other variables into the research. By testing a participants existing products, used for tasks they regularly carry out and in their own environment, we should obtain an accurate indication, which aspects of normal operation cause major usability issues in normal use. The research methodology involved detailed questioning and observation of a relatively small number of participants, 58 in total. The reason for this is that in order to identify the key usability issues a researcher will not only need to ask the opinion of the participant but also observe where problems occur, record events and

[2] This work has been partly funded by the European Commission through ICT project VICON (Virtual User Concept for Inclusive Design of Consumer Products and User Interfaces). The authors wish to acknowledge the European Commission for their support.

encourage greater feedback from the user. Initially it had been suggested that questionnaires could be used as part of this study. They could have given the research team results from a larger user group but the depth and quality of information collected was considered to have been much lower and thus less appropriate to this specific research task.

It was considered important that the research should be carried out in suitable environments. For a washing machine this should be in the user's home as it introduced usability issues related to lighting, ambient noise, heating and the availability of space to access and use the product. When using a mobile phone the environment makes a considerable difference to the usability issues encountered when using the product. For practical reasons it was decided that each user should carry out tasks using their mobile phone in their normal domestic environment. However, whenever possible it was suggested that the user were observed using the phone in both low and high lighting conditions. Additionally the users were observed using the product in both a static and mobile environment, so the users were encouraged to use their mobile phone both indoors and outside.

We looked that the user groups comprise each 10-15 people of hearing impaired participants, visually impaired participants, manual dexterity impaired participants, and elderly participants exhibiting low levels of visual, hearing and manual dexterity impairment. Each of the three research partners (RNID, NCBI and FIT) recruited 20-30 participants for this research, the majority of whom are likely to be elderly. Most of these people had multiple mild to medium impairments but out of the total number (60-90) each 10-15 had a more significant hearing loss, a more significant sight loss and more significant problems with manual dexterity. In selecting participants we needed to screen for people with cognitive issues which might affect their ability to use a mobile phone or washing machine. Those with severe cognitive problems were excluded from the research. It was further checked that the participants regularly use the specific product-types and determined what kind they currently use. - The main observational research methodology was mainly carried out by RNID with collaboration from the other research partners.

The researcher directed the participant to carry out specific tasks related to the everyday use of the products, made objective observations and asked relevant questions. This procedure followed a standard questionnaire/methodology formulated before and during the pilot research. These major tasks are all related to the normal use of a product and aim to carry out specific functions in the most logical or practical order. Typically these major tasks were to ask the user to perform certain typical tasks in a logical order, or observe what the user actually does and any usability issues related to carrying out the task. This will involve recording the steps taken, usability issues and any obvious mistakes/successes (for example spilling washing detergent). Furthermore users were asked how easy/difficult they found the task. This had to be explored in detail, including talking through the process, if there were particular problems or if it was deemed relevant. The observer needed to investigate how much each usability issue was down to the specific impairment/s of that user as opposed to being more specific to product design or environmental factors. Observations were recorded in written, abbreviated form.

6 Results

In fig. 2 we give the personal information of the 58 subjects selected an in fig. 3 their impairments. The complete study is available online [3].

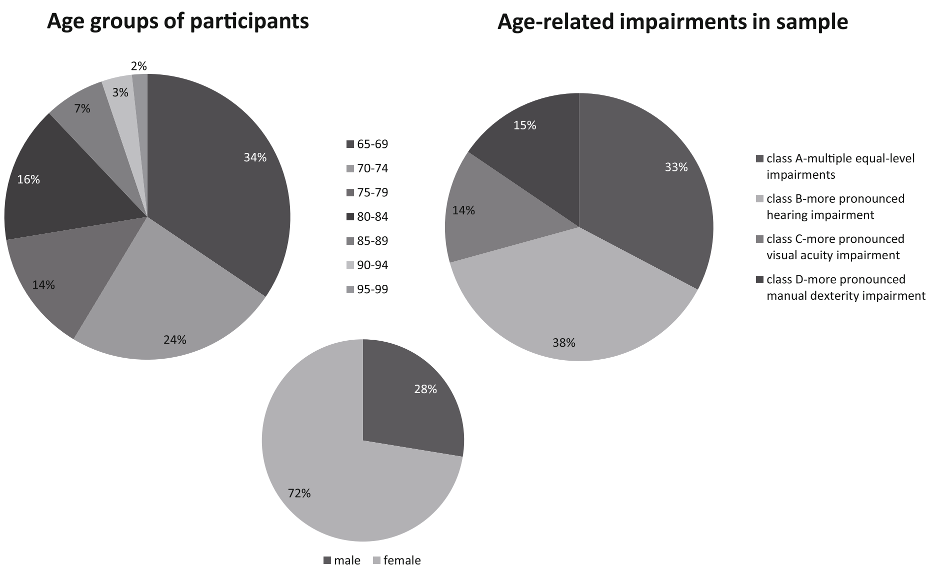


Fig. 2. Personal information (gender, age and impairment)

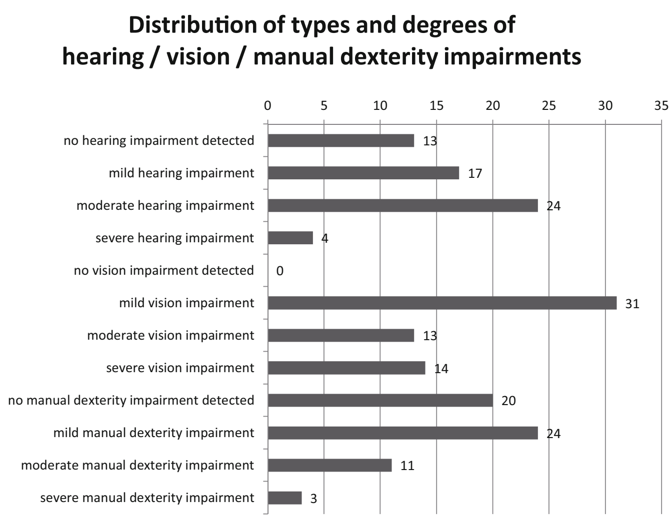


Fig. 3. Distribution and types of the different user impairments

7 Results for the Washing Machine

There were a small number of problems related to using the door catch. In these cases either the catch was too small, a dual-action was required to operate the catch or a push button control required force to operate. There were a small number of problems (n=4) related to unloading the washing machine. 2 participants (user groups A and D) found it hard to remove heavy, wet laundry from the drum. Participants generally found the detergent dispenser easy to use; this included opening and closing the dispenser, adding detergent and conditioner. 6 participants had problems removing the detergent dispenser. For 2 users (both from group A) difficulty and effort were given as their reasons, another group A user commented that it was ‘finicky’ and a user from group D found it ‘awkward’. The majority of participants had no problems in setting wash programmes; 94% (44 out of 47) were able to set programme 1 (their most commonly used wash programme) correctly in one attempt and 95% (38 out of 40) were able to set programme 2 (a second wash programme chosen by the researcher) correctly on the first attempt. Users were asked to operate the different types of minor control. Of the 46 washing machines included in this research; 40 had at least one type of minor control, 25 had two and 18 had three. Although most of the products had a wide range of features and minor controls, in most cases the design of these minor controls was identical - presumably for aesthetic reasons. It is common for a washing machine to have a row of between 3 and 6 identical minor controls between the main programme control and the start or on/off control – these minor controls only differing in their labeling. 8 participants had problems when using the programme guide; 4 (2 from user group A and 2 from B) had to bend or kneel down to read it, 3 (1 each from groups A, B and C) said it wasn’t easy to read and single users commented that they would prefer to use the manual instead (group A), the guide was too small (group C) and the language was not easy to understand (group B). We wanted to see if people had any problems in identifying when their wash had ended, but only 2 users had problems with this. 1 complained that there was no signal to alert the user and the other was confused by an indicator light indicating ‘crease protect / end’. Observations were either positive or simply described how the user became aware that their wash had finished. Many participants (n=18) washing machines had a light indicating ‘finished’ or ‘end’ and almost as many (n=17) relied on the fact that the final spin is very noisy and they simply wait for this to end. Similarly 9 users commented”it stops” which probably also relates to the loud final spin coming to an end. Other users relied upon the position of the main (electro-mechanical) programme control, the timer being at ‘0’, an audible bleep or the display reading ‘end’. Interestingly, 4 users actually timed their wash so they already had a good idea when it will have ended. Only 36% of the participants ever clean the filter on their washing machine. When asked to describe how they do it 6 said it was easy to do and 4 users had self-cleaning models without a filter. Other users reported problems with having to bend, having to use a shallow tray to collect any water spillages and the fact that it could be a difficult and fiddly task to carry out. So, the reasons why most of the users don’t clean the filter could be due to these difficulties or maybe even because they aren’t aware that it needs to be cleaned regularly.

8 Results for the Mobile Phone

Many users (n = 15) reported having difficulties using the On/Off control. One issue was related to the force required to press the button. Many users (n = 10) had difficulty with this task either as a result of having to use too much force or because they experienced pain or discomfort. 1 user reported leaving the phone on continuously, to avoid the difficulty of turning it on and off, as she had arthritis. In the observations 12 people reported that the button required force to operate, so this was obviously a significant problem.

8 participants had problems when making a voice call and all of these problems were related to the operation of the number keys and other controls. 3 users had problems due to the number keys being too close together, so they often pushed more than one button at the same time. Other individual problems included buttons being too small and fiddly, buttons being difficult to operate if the user has long finger nails, problems deleting incorrect numbers, the numbers on the keys being hard to read and force being required to operate the keys. No users reported having problems when receiving a voice call. Observations recorded few negatives but in 3 cases the ring was too quiet and 1 user had problems with the keypad lock. There were many positive observations made including; loud ring, strong vibration, easy to know when I am receiving a call and screen lights up. Over half (n =25) of the users used the ‘Ring and Vibrate’ setting to alert them to incoming calls, but almost as many (n=19) used ring only. 2 users were alerted to calls by ring and light but none chose to be alerted by vibration alone. When asked why they chose a particular alert, 10 users said that that was the way the phone was set up for them - the phone was either already on this setting, or a family member had selected it for them. Other replies generally explained and justified most users chosen method of being alerted to a call.

When asked if it was easy to hear the speech during a call, 6 users reported having problems. For 4 participants (2 each from groups B and C) their phone was too quiet, 2 users (1 each from groups B and C) said the speech didn’t sound clear and 1 user (group A) could only hear the odd word. When asked if it was easy to understand the speech during the call, 9 users reported having problems. 6 participants (2 each from groups B and C) couldn’t understand the recorded message and said it wasn’t clear. 2 users (1 each from groups A and B) found the accent of the voice difficult to understand, 2 others (both from group B) found the speech too fast and 2 (one each from groups B and C) found that the low volume made it hard to understand the speech. Participants were asked to alter the volume of the recorded speech message during a call and 10 people had problems carrying out this task. 6 users (3 each from groups A and B) didn’t know how to alter the volume, but the researchers suspect that many others were unsure about what to do and were only vaguely aware that the volume could actually be changed.

Of the 49 participants who took part in this research, only 26 (59%) send SMS text messages and answered questions on these tasks. This indicates that although many elderly people now have mobile phones, their primary use is likely to be for occasional voice calls instead of text messaging. 28 participants attempted this task. Similar to the results for receiving a voice call, 50% (n =14) used the ‘Ring and Vibrate’ setting to alert them to an incoming message and 43% (n=12) used Ring/Tone only. Many users (n=17) had problems when they were asked if the number keys were large enough for them. 10 users specified that the keys were too small; 7 said that they found the keys too small and fiddly, 2 stated that they tended to push two keys at once and 2 obviously found the keys too small. Other problems recorded include the number keys not being arranged in a straight line, the buttons not protruding enough, having to use fingernails to operate the small controls and a rounded button shape making it too easy for the finger to slip off and press the neighbouring button by mistake.

Some users (n = 10) reported not finding the display easy to read or having problems with it. 3 users (2 from group B and 1 from C) found the displays on their phones too small and 1 of these also disliked having dark grey figures on an orange background. 2 users (from group A and B) found that their phone had an energy saving function which darkened the screen, this happened too quickly for them making the display difficult to read. A single user (group A) commented that the calendar and menu functions looked quite faint.

For mobile phones, this was the usability issue which had the most problems associated with it. 21 participants reported having problems when they were asked 'do you understand the icons or descriptions'. 12 users said they were not sure about some of the words or icons used and 3 people did not understand the menu functions and so don't use them. 2 users thought the descriptions were not intuitive and another 2 thought the instructions and language were too complicated. 2 users commented that they liked the clear diagrams (icons) and words.

Less than half of the users (n=23) attempted to add contact details to the phonebook.

9 Discussion of Results

No significant differences were found between the results for each of the user groups. Therefore the results have been combined to indicate the main usability problems encountered by 'typical elderly people'. The main problems which the user group encountered were with washing machines when removing and replacing the detergent dispenser, cleaning the detergent dispenser, cancelling a wash program, labelling of the controls and cleaning the filter. The main problems with mobile phones were encountered with turning the phone on or off , making a voice call (specifically related to using the controls), hearing and understanding speech during a call, adjusting the volume, using the main number controls, using the additional controls, understanding and reading the labelling, understanding the features and menu functions, and the keypad lock function.

Users were asked to use the Doro Easy®410gsm phone to see how this specially designed product compared with their existing phone. The results showed that this phone proved to be a significant improvement over normal, mainstream products and that; as a result, most of the problems which users reported on their mobile phones can be solved by following inclusive design principles and by focusing ones efforts upon the needs of people who might have mild to medium impairments. What's more, the Doro phone is proving to be commercially viable.

Other findings of the research included: Many elderly users did not use SMS text messaging nor add contact details to the phonebook on their mobile phone. Elderly users did not have problems recharging their mobile phones. Many were well prepared for this regular, but potentially tricky, task and had developed it into a routine.

10 Conclusion

In this paper we present first results from an ethnographic study on elderly people with mild to medium impairments on using everyday products. The results will be beside others a source of a knowledge management system for a Virtual Laboratory using Virtual User Models. It will allow an extension to existing design processes. The knowledge on the needs of people with mild to medium age related impairments will in this way be provided directly to the designer. The entire design process of UIs of consumer products will be covered. We expect to further approach our aim of inclusion to as far as possible make products universal, with users both with and without disabilities having similar choices of UIs to suit circumstance and preferences.

References

1. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human Computer Interaction. John Wiley and Sons, UK (2002)
2. ICF International Classification of Functioning, Disability and Health, WHO (2001), `http://whqlibdoc.who.int/publications/2001/9241545429.pdf` (accessed October 18, 2010)
3. `http://www.vicon-project.eu` (accessed February 1, 2011)

Effects of Age Groups and Distortion Types on Text-Based CAPTCHA Tasks

Chih-Hsiang Hsu and Ying-Lien Lee*

Department of Industrial Engineering and Management, Chaoyang University of Technology, Taichung County, 413, Taiwan
Tel.: +886-4-23323000; Fax: +886-4-23742327
No. 168, Jifong E. Rd., Wufong Township, Taichung County, 413, Taiwan
yinglienlee@gmail.com

Abstract. Completely Automatic Public Turning test to tell Computers and Humans Apart, or CAPTCHA, is a security measure that guards a system from exploitation by the discrimination between a real human being and an automated computer program via the method of presenting to the unknown user the challenges that are hard for computer yet easy for human. Focusing on text-based CAPTCHA, this study conducted an experiment to study the effect of age groups and distortion types on the CAPTCHA task. Twenty-four participants were recruited to take part in the experiment, where twelve of them were in the senior group and twelve in the young group. Participants were observed to use three general steps: recognition, rehearsal, and motor response. With the inevitability of the security measure and the increasing population of senior netizens, this study has important implications for the design of CAPTCHA systems.

Keywords: CAPTCHA; age group; distortion.

1 Introduction

CAPTCHA, the acronym of "Completely Automatic Public Turning test to tell Computers and Humans Apart", is a security measures that guards internet services against automated exploitations with abusive purposes [1]. The merit of a CAPTCHA system lies in the system's capability to tell whether the user in question is a real human being or a robot program. A CAPTCHA process typically involves a session in which computer-generated questions are presented to users whose true identities are unknown to the system. Based upon the answers replied by the users, the CAPTCHA system determines whether the user in question is a human or not. To effectively tell computers and humans apart, the proposed questions have to be hard for computer to solve, yet easy for human to answer [2].

Among the techniques of breaking text-based CAPTCHA protection mechanism, Optical Character Recognition (OCR) programs is one of the commonly deployed

* Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 453–455, 2011.

methods to defeat the protection [3]. Generally, OCR programs recognize characters contained in an image via three steps [4]: (1) pre-processing of the image to make the image suitable for further processing, (2) segmenting the image into regions in which each region contains only one character, and (3) identifying the character in each region. To lower the success rate of character recognition by the OCR programs, CAPTCHA systems usually distort the images in certain ways to complicate the steps OCR programs typically employ. For example, a CAPTCHA system may warp the image or add extra elements (such as lines of various slopes, or circles of various radii) to counter the attacks of OCR programs. With the distortion, however, CAPTCHA questions may become difficult to solve even for human being, which may stress the user's cognitive system and vision system, and make a potential customer walk away [2, 5].

2 Method

Two groups of participants were recruited to take part in the experiment. The twelve participants in the senior group were between 50 to 56 years old with an average of 53.2 years, while the other twelve in the young group were between 23 to 24 years old with an average of 23.6 years. All participants had normal or corrected to normal visual acuity, and had normal color vision.

The experiment was a two-by-six mixed factorial design in which the two factors were age group (encoded as AGE) and CAPTCHA type (encoded as TYPE). The factor AGE was a between-subject factor with two levels, young and senior groups. The factor TYPE was a within-subject factor with six levels, each representing a kind of common CAPTCHA distortion.

The experiment was conducted by one experimenter in the settings mentioned above. A consent form with brief description of the study was given to each participant before the experiment could begin. Once the form was signed, the experimenter introduced the procedure to the participant and helped the participant familiarize oneself with the tasks involved. As described before, a participant had to finish six sessions to conclude an experiment. The first session was always the reference type (Type I), while the types of the latter five sessions were randomly determined for each participants. The participant could practice the kind of CAPTCHA stimuli in a practice session before a formal session began. Stimuli used in the practice and formal sessions underwent the same kind of distortion, but the texts of the two were different.

3 Results

Table 1 summarizes the means and standard deviations (enclosed in parentheses) of the dependent variables for distortion types and age groups (senior and young). The results of non-parametric analysis of the data revealed that the participants of two age groups differed significantly in terms of response time, error rate, and NASA-TLX score. Distortion type had significant effect on response time, error rate, CFF (Critical flicker fusion), and NASA-TLX score. Post-hoc analysis showed that Blot Mask and Line Mask were the hardest CAPTCHAs, while Thread Noise, Global Warp, and Geometry

Noise were on a par with Normal Type (no distortion). Dependent variables were also correlated to each other.

Table 1. Descriptive statistics of the dependent variables by types and age groups

	Mean (SD)	Response time in ms	Error rate	CFF in Hz	NASA-TLX score
Normal	Senior	6694.56 (3177.19)	0.04 (0.04)	0.67 (0.69)	16.25 (2.14)
Text	Young	2755.03 (1131.92)	0.02 (0.03)	0.46 (0.4)	14.75 (3.49)
Blot	Senior	9960.46 (3645.76)	0.13 (0.06)	2.38 (0.61)	19.67 (2.31)
Mask	Young	4664.66 (1366.68)	0.07 (0.04)	2 (0.52)	17.25 (2.8)
Line	Senior	11974.85 (4385.73)	0.22 (0.06)	2.67 (0.39)	20.67 (2.02)
Mask	Young	6733.49 (2348.31)	0.23 (0.08)	2.63 (0.8)	19.08 (2.75)
Thread	Senior	6158.93 (2626.14)	0.05 (0.02)	1.29 (0.69)	16.92 (2.11)
Noise	Young	3072.86 (1358.78)	0.04 (0.03)	1 (0.74)	15.25 (3.05)
Global	Senior	6261.75 (2394.85)	0.06 (0.04)	1.54 (0.4)	17.25 (2.22)
Warp	Young	2707.77 (962.26)	0.02 (0.01)	1.33 (0.49)	14.92 (2.87)
Geometry	Senior	6104.62 (2082.41)	0.05 (0.04)	1.21 (0.58)	17.17 (2.69)
Noise	Young	2692.46 (1167.67)	0.01 (0.01)	0.96 (0.62)	14.83 (2.72)

4 Discussion and Conclusion

The results of the present study verified that participants of different age groups differ significantly in terms of response time, error rate, visual fatigue, and workload. In addition, some types of CAPTCHA design have significant impacts on the dependent variables. When designing a CAPTCHA system, one should take into account the basic steps users take during the interaction with such system. A CAPTCHA system may be so designed that the recognition step is as easy as an undistorted one for human users. Yet, not only distortion technique, but also the character set, that can impact the recognition and the rehearsal steps.

References

1. von Ahn, L., et al.: CAPTCHA: Using Hard AI Problems for Security. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 294–311. Springer, Heidelberg (2003)
2. Chellapilla, K., et al.: Designing human friendly human interaction proofs. Presented at the ACM CHI (2005)
3. Kolupaev, A., Ogijenko, J.: CAPTCHAs: Humans vs. Bots. IEEE Security & Privacy 6, 68–70 (2008)
4. Chellapilla, K., et al.: Building Segmentation Based Human-friendly Human Interaction Proofs. Presented at the 2nd Int'l Workshop on Human Interaction Proofs (2005)
5. Elson, J., et al.: Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization. Presented at the 14th ACM Conference on Computer and Communication Security, Alexandria, VA, USA (2007)

Evaluating Usability of Web-Based Electronic Government: Users' Perspective

Zhao Huang and Laurence Brooks

School of Information System, Computing and Mathematics, Brunel University, Uxbridge, Middlesex UB8 3PH, UK
{zhao.huang,Laurence.brooks}@brunel.ac.uk

Abstract. Electronic government (e-government) has been rapidly developed to increase services delivery and users' access to government information. Users' interaction with e-government is largely dependent upon how easy to use e-government websites. In this aspect, usability needs to be importantly addressed when developing e-government. As such, an empirical study is conducted to evaluate usability of current e-government websites focusing on users' perception and performance. The results indicate that a number of usability problems have been found in the target e-government websites. Furthermore, performance results show the different levels of users' interaction with the e-government websites evaluated. More significantly, a close relationship between users' perception of usability and their performance is indicated. These findings can help designers identify users' usability requirements and draw their particular attention to further develop more usable e-government websites.

Keywords: Web-based e-government, Usability, Users' perception, Users' performance.

1 Introduction

Traditional government is a complex and mammoth bureaucracy, which makes access to information difficult and provision of government services frustrating for users [1]. Users want to establish a new approach for better government, which enables easier access, richer information resources, higher quality services and more enjoyable participation [2]. Furthermore, government is required to change the way of administering and processing official business [3], delivering government services and generating greater efficiency for all participation [4]. In response to these requirements, e-government becomes an outstanding solution [5].

Today, governments worldwide have rapidly developed their web-based e-government systems. Among the 192 member countries of the United Nations, nearly 98% of countries have built web-based e-government systems [6]. However, the average e-government website usage by users is low, about 30% [4]. Among the reasons for this, usability has been found to be a major factor influencing users' engagement with e-government websites. A higher level of usability can lead to better systems quality and help users more easily and effectively accomplish what they want

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 456–465, 2011.

to do on the e-government websites. As such, usability of e-government websites needs to be considered when developing e-government websites.

However, current research has not paid enough attention to evaluating usability of e-government websites. Without addressing usability at the detailed level in e-government websites design, e-government will not be fully used by the wider range of users. Since e-government is used by diverse users who have heterogeneous backgrounds, there are a range of requirements for usability from e-government. Such different requirements make usability identification difficult for designers trying to develop more usable e-government. To this end, an empirical study is conducted to evaluate the usability of current e-government websites. By doing so, it can provide insight into e-government usability. In addition, this study addresses the users' perspective, which increases understanding of users and their usability needs. This offers concrete prescriptions for developing more user-centered e-government websites that can support users achieving the desired services outcomes and so generate greater participation. To implement the evaluation, the heuristic evaluation method is considered as appropriate, with its usefulness and applicability already proven in many studies [7]. In detail, the heuristic evaluation is based on users' perception of Nielsen's set of usability heuristics. Furthermore, in order to obtain comprehensive evaluation, users' performance is also measured showing the level of users' interaction with the e-government websites when they perform specific tasks.

The paper is structured as follows: section 2 reviews relevant studies to indicate the importance of usability to e-government websites. This is followed by designing an empirical study to evaluate usability (section 3). Section 4 discusses the evaluation results. Finally, the conclusion and future research are suggested in section 5.

2 Relevant Studies

2.1 E-Government Websites

Since all e-government information and services are delivered through e-government websites, these websites can therefore be seen as the interface of the e-government, serving as a window for users to communicate with governments [8]. Alongside e-government development, e-government website functionality has significantly evolved. Initially, the website is simply used to publish information. However, the website matures quickly and functionality is increased by adding search facility, personalized web content and online transaction [9]. Recently, website functionality is underlying systems integration [8]. With high levels of integration, users can simply follow a single registration process to become involved in multiple online government services [10]. An e-government website has the potential to change the way that users access and interact with government. As such, e-government websites are a key priority when governments develop their e-government systems and create electronic relationships [11]. In the UK, 30% of e-government projects are focused on website development [12]. However, the target of great users' interaction with e-government remains a challenge. Evidence indicates that usability is a vital factor in deciding users' participation.

2.2 Usability

Usability can be seen as a measurement related to how useful and user-friendly the system is. In detail, usefulness is the degree to which users think that using the particular system can improve their performance [4], while user-friendliness is the perception of aesthetic design regarding interface features [13]. Usability is an important factor in system quality. While a successful and preferred website generally refers to one with high usability, equally, usability can significantly influence users' preferences, opinions and attitudes [14]. Furthermore, usability has big impacts on user performance with systems. A study investigating users' expectations about e-government shows that users' preferences closely relate to usability in terms of the degree of access of e-government services, findability of the e-government website, loading speed of the pages, the usefulness of information provided on the site and flexibility that is being offered through the e-government website [15]. A high level of usability enhances users' expectations, which has a decisive effect on their use of e-government. Thus, there is a need to evaluate usability of e-government websites to improve its usage.

3 Methodology

An experimental study is carried out to evaluate the usability of current e-government websites. There are three research instruments: the selected e-government websites, the task sheet and the usability questionnaire. The e-government website is selected as representative of e-government. The task sheet contains a set of tasks that is developed for the participants to perform. The usability questionnaire is designed to identify the participants' perception of usability.

3.1 E-Government Websites Selection

The local level of e-government website is selected in this study, because a) it is the closest level for users; b) it is frequently used by the public since the local level is more informational for users and focuses on the needs of users in accessing information and services [16]; c) it can show the effects of e-government on users [17] and d) evidence from studies finds that bigger challenges exist at the local level of e-governments and in their website design. Therefore, the study uses three local e-government websites in the U.K: called London Authority 1, 2 and 3.

3.2 Task Sheet

The participants are required to perform a set of practical tasks on the target e-government websites. Such tasks are representative activities that users would be expected to carry out on an e-government website. The task sheet is used to detail these tasks for the participants for the usability evaluation. Generally, there are three categories of e-government services [11]: information dissemination, products and services offered and user participation. Information dissemination is related to the provision of all types of government information. Products and services offered refers to delivering one-way services, such as document downloads. User participation involves users interacting with two-way services, for example, tax payment. Based on these service categories, a set of relevant task were designed.

3.3 Usability Questionnaire

A questionnaire is used to identify users' perception of usability of the target e-government websites. The design of this questionnaire is based on Nielsen's usability heuristics [18]. There are three steps in the questionnaire design. First, extend the existing usability heuristics to fit the specific needs of e-government. Second, a set of associated criteria for each heuristic are developed in order to focus on the detailed aspects of usability. Finally, the specific questions are developed, based on these heuristic criteria.

Extension of usability heuristics. Nielsen's usability heuristics (Tables 1) have been widely used for usability inspection and its applicability and validation have been proved in many studies [7]. In particular, these heuristics can effectively discover usability issues in relation to website design [19]. As such, this study uses these heuristics as a starting point to evaluate the e-government websites usability.

Table 1. Nielsen's usability heuristics [18]

Usability Heuristics	Explanations
1. Visibility of system	The site should keep users informed about system status.
2. Match real world	The site should use the user' language, real-world conventions.
3. User control	The site should make undo and redo functions available.
4. Consistency	The site should keep the same design features through the site.
5. Error prevention	The site should support users to overcome errors.
6. Recognition rather than recall	The site should make objects easy to remember. In addition, instruments on the site should be visible and easily retrievable.
7. Efficiency of use	The site should consider usage for novice, experienced users.
8. Aesthetic design	Dialogues should not contain irrelevant information.
9. Recover errors	The site should indicate the problem and suggest a solution.
10. Help and documentation	The site should provide help and documentation that can be easy to search, focus on the users' tasks, list concrete steps.

However, these heuristics were developed many years ago and use for general website evaluation purposes. To meet the needs of e-government websites, it is necessary to extend Nielsen's heuristics. Relevant studies indicate that e-government is used by a wide range of people, while interoperability is important in terms of information and service exchange [11]. In addition, since e-government is used by diverse users who have heterogeneous skills, therefore, e-government should support these users to complete services. Furthermore, during users' interaction, e-government should respect their users at all times [20]. Therefore, the existing Nielsen's heuristics are extended by adding three further heuristics: 'Interoperability', 'Support users' skills' and 'Pleasurable and respectful interaction' (Table 2).

Table 2. Extended usability heuristics

Extended Heuristics	Explanations
11. Interoperability	The site should make all service parts, design elements, the site functions work as a whole to support user task completion.
12. Support users' skills	The site should support, extend users' current skills.
13. Respectful interaction	The site should present a pleasant design, treat with respect.

Associated criteria development. Although Nielsen's usability heuristics are extended, these heuristics are still too general to develop the usability questionnaire. It may result in a usability evaluation without enough depth. Furthermore, the lack of detailed analysis may lead to failure in specific usability problem detection. Accordingly, it is important to develop a set of associated criteria. Such criteria are developed from relevant usability studies and the interpretation of wider e-government studies. The findings of these studies are used to identify which website design features may influence users' perception of usability or cause problems when users interact with systems. These features are extracted to develop the criteria and then grouped into corresponding heuristics.

Usability questionnaire design. Based on the associated criteria, a questionnaire is created for the purpose of capturing the participants' assessment of usability of the target e-government websites. The participants are required to respond using a five-point Likert scale, which can indicate the participants' agreement level to the statements.

3.4 Evaluation Procedure

To conduct the evaluation, 36 participants were assigned to evaluate three target e-government websites. Each target e-government website evaluation involved 12 participants. Each participant follows the same evaluation process, which are: free review, task implementation and completing the questionnaire. Free review allows the participants to look through the target e-government website several times. They can freely either look at the overall e-government website or focus on the specific website design elements. Subsequently, the participants are required to complete a set of tasks on the e-government website. Having accomplished all the tasks, the participants are finally asked to fill in the usability questionnaire.

3.5 Data Analysis

The data analysis techniques used are the one-way ANOVA and the one-sample T-test. Statistical analysis is conducted using SPSS for windows (version 13). The significant value (P) is pre-defined as less than 0.05. In detail, to indicate whether the three London Authorities have a difference in the participants' overall perception of usability, a one-way ANOVA is conducted with three London Authorities as independent variables and usability perception as the dependent variable. Similarly, the one-way ANOVA is employed to indicate whether the three London Authorities show differences in users' performance. In addition, to indicate whether there is a difference between the perception of overall usability and the perception of specific usability features in each target London Authority, a one-sample T-test is conducted.

4 Discussion of the Results

4.1 Users' Perception

In general, the results indicate that usability has been considered in the three e-government websites as a number of usability strengths have been detected (Table 3).

However, a number of usability problems have also been identified in each target e-government website (see Table 4). Among these problems, a lower mean score indicates a more serious problem. In London Authority 1, the most serious usability problem is that users are confused by links that have many different colors. Link color is used to indicate different resources within the site. An appropriate number of link colors can visually support users to distinguish between resource differences, so that target information can be easily located to meet users' needs. Using this, users with limited color vision can quickly recognize the difference among subjects [21]. On the contrary, links with many different colors may visually influence the site's appearance and obstruct users' color vision, which may result in difficulty and confusion in information identification in London Authority 1.

In London Authority 2, the most serious problem found is that the options on the home page are not clearly enough presented. Users' subject recognition would be better supported by having clearly presented more understandable options. It can simplify the content presentation and improve its readability. As such, it can help users to quickly understand the subject presented on pages and easily select the relevant options to obtain their expected information. However, options without clear presentation can lead to page content complexity. A starting page with high complexity makes user less pleasurable and users perform worse on search and recognition subjects on such pages [22]. Therefore, the problem of the subject options not being clearly presented on the home page of London Authority 2 affects subject content presentation, which may cause users difficulty with information seeking.

Regarding London Authority 3, the most serious usability problem is that subject categories are presented without a logical order. A logical order of subjects is used to indicate a sequence of information organization, which supports users having a sensible way to scan subject information. It assists users' understanding of the overall subject arrangement and reduces memory load problems. When topics are arranged with a particular order, users are able to easily locate items; remember items of interest viewed previously and access primary information quickly [23]. Conversely, when subject categories are presented without a logical order, users may feel it difficult to scan and find subjects among the categories on the site.

Table 5 presents the overall usability evaluation of the three e-government websites. The results indicate that London Authority 2 has the worst usability assessment, London Authority 1 is next, while the e-government website with best overall usability assessment is shown to be London Authority 3.

4.2 Users' Performance

Table 6 presents the participants' performance with the three London Authorities. Users' performance is measured by a number of performance criteria through

Table 3. Usability strengths in London Authorities 1, 2 and 3

London Authority 1 usability strengths	Mean (SD)
Users can easily move forward and backward within the site.	4.33 (0.49)
Significance	T=6.215, P=0.000
Site offers "A-Z" service that supports users to quick find the relevant information for the specific tasks.	4.08 (0.79)
Significance	T=2.767, P=0.018
Different displays on each page are compatible through the site	3.92 (0.52)
Significance	T=3.139, P=0.009
London Authority 2 usability strengths	Mean (SD)
The options/links used are all working properly.	4.25(0.452)
Significance	T=7.123, P=0.000
Each image corresponds to each context.	4.17 (0.577)
Significance	T=5.080, P=0.000
Users can easily move forward and backward within the site.	4.00 (0.853)
Significance	T=2.762, P=0.018
A title with every page clearly indicates the subject of the content.	3.92 (0.793)
Significance	T=2.607, P=0.024
The site's functionality supports users to complete most tasks.	3.75 (0.452)
Significance	T=3.294, P=0.007
London Authority 3 usability strengths	Mean (SD)
Each page is always followed the same display format.	4.67 (0.492)
Significance	T=5.816, P=0.000
The site' functionality supports users to complete most tasks.	4.67(0.492)
Significance	T=5.816, P=0.000
A title on every page clearly indicates the subject of the content.	4.58 (0.669)
Significance	T=3.852, P=0.003
Key information/subject is placed in a central location on the page.	4.50 (0.674)
Significance	T=3.391, P=0.006
It is easy to operate the e-government website.	4.42 (0.669)
Significance	T=2.988, P=0.012
It is quick to change the particular data in a previous section so users do not need to retype all the data when they go back.	4.42 (0.793)
Significance	T=2.519, P=0.029
Users can move forward and backward within the site.	4.25 (0.622)
Significance	T=2.285, P=0.043

(SD= Std. Deviation)

observation. These criteria include amount of online help required, time spent completing all tasks; number of steps to finish tasks and number of successful tasks completed. The results show that the participants in London Authority 2 required more online help and took more steps to accomplish tasks than those in London Authorities 1 and 3. Furthermore, regarding the number of successful tasks completion, the participants in London Authority 2 finished fewer tasks than the participants in London Authorities 1 and 3. These are reflected in the findings of the overall users' perception of usability, which shows that London Authority 2 has the worst overall usability of the three target e-government websites. This implies that the overall users' perception of usability positively influence users' performance.

Table 4. Usability problems in London Authorities 1, 2 and 3

London Authority 1 usability problems	Mean (SD)
Users are confused by links that have many different colors.	2.32 (1.084)
Significance	T=-3.303, P=0.007
Online help function is not clearly indicated on the website.	2.33 (1.155)
Significance	T=-3.350, P=0.006
It is difficult to switch between online help and current work.	2.75 (0.866)
Significance	T=-2.800, P=0.017
London Authority 2 usability problems	**Mean (SD)**
Some options on the home page are not clearly presented.	2.17 (1.030)
Significance	T=-3.879, P=0.003
Users are confused by links that have many different colors.	2.25 (0.866)
Significance	T=-4.280, P=0.001
The site sometimes does not indicate a task's progress.	2.33 (0.888)
Significance	T=-3.851, P=0.003
Links already visited are not clearly marked.	2.50 (1.243)
Significance	T=-2.285, P=0.043
The site allows users to skip over the order of the process.	2.67 (0.778)
Significance	T=-2.907, P=0.014
London Authority 3 usability problems	**Mean (SD)**
Subject categories are presented without a logical order.	2.83 (1.030)
Significance	T=-3.386, P=0.006
Users are confused by links that have many different colors.	2.58 (0.669)
Significance	T=-6.511, P=0.000
Links already visited are not clearly marked.	2.92 (1.084)
Significance	T=-2.952, P=0.013
Information is unbalanced between breadth and depth.	3.00 (0.853)
Significance	T=-3.412, P=0.006

(SD= Std. Deviation)

Table 5. Overall usability assessment in the target London Authorities

	London Authority 1		London Authority 2		London Authority 3	
	Mean	SD	Mean	SD	Mean	SD
Overall usability	3.445	0.304	3.323	0.367	3.843	0.275
Significance						F=8.784, p=0.001

(SD= Std. Deviation)

However, regarding performance in terms of time spent for all tasks completion, the results indicate that the participants who are in London Authority 1 spent more time to finish all tasks than those who are in London Authorities 2 and 3. Such results are not echoed in the findings of the overall users' perception of usability, in which London Authority 1 has not the worst overall usability among the target e-government websites. Based on the results, a possible explanation is that the participants' performance may be not only affected by the overall perception of usability, but also affected by the particular perception of usability. As the most serious usability problem identified in London Authority is that users are confused by links having many different colors. This problem closely relates to the heuristic “aesthetic design” (Table 1). Evidence from relevant studies indicates that system aesthetics can be seen as apparent usability, which is perceived more quickly than other attributes of usability [24]. Therefore, users’ perception of website aesthetic design may also influences users’ performance, especially in London Authority 1.

Table 6. Users’ performance in London Authorities 1, 2 and 3

	London Authority 1		London Authority 2		London Authority 3	
	Mean	SD	Mean	SD	Mean	SD
Online help required	0.250	0.452	0.583	0.669	0.000	0.000
Significance					F=4.733, P=0.016	
Number of steps	60.417	13.104	81.833	20.687	50.167	16.297
Significance					F=10.862, P=0.000	
Tasks completion	1.139	0.117	1.148	0.086	1.065	0.088
Significance					F=2.590, P=0.090	
Time spent	26.627	8.905	21.721	8.579	16.209	8.102
Significance					F=4.474, P=0.019	

(SD= Std. Deviation)

5 Conclusions

This study has evaluated usability of current e-government websites in the UK. A number of usability problems have been found in the target e-governments. This suggests that current e-governments need to improve their usability. In addition, it appears that usability issues have not been considered at a detailed level in e-government websites design, which influences users’ task performance. These results can help designers understand e-government websites usability, especially identifying users’ usability requirements. Furthermore, these identified usability problems can draw designers’ particular attention to their websites and support them to further improve usability. Additionally, the common usability problems detected can be used as reference to check usability for other e-government websites. However, this study has some limitations. For example, regarding usability criteria development, some criteria might be found to relate to more than one heuristic; however, the study grouped these criteria into one heuristic based on their key features. Moreover, to fulfill a thorough usability study, there is a need to offer concrete prescription for the identified usability problems in order to improve usability of the target e-government websites. Future research will propose the design solutions and examine the effects of the proposed design solutions on the usability problems identified in this study.

References

1 Cairns, G., Wright, G., Bradfield, R., Heijden, K., Burt, G.: Exploring e-government futures through the application of scenario planning. Technological Forecasting and Social Change 71(3), 217–238 (2004)

2 OECD: OECD E-Government Studies: the e-government imperative, Paris (2003)

3 Yang, L., Lu, Y., Fu, G.: Study on e-government construction. In: Proceeding of the 4th International Conference on Entertainment Computing, pp. 542–548 (2005)

4 Kumar, V., Mukerji, B., Butt, I., Persaud, A.: Factors for successful e-government adoption: a conceptual framework. Electronic Journal of E-government 5(1), 63–76 (2007)

5 Kelly, E.P., Tastle, W.J.: E-government and the judicial system: online access to case information. International Journal of Electronic Government 1(2), 166–178 (2004)

6 UN Public Administration Programme 2010: United Nations e-government survey (2010), `http://www2.unpan.org/egovkb/global_reports/10report.htm`

7 Hvannberg, E.T., Law, E.L., Larusdottir, M.K.: Heuristic evaluation: comparing ways of finding and reporting usability problems. Interacting with Computers 19, 225–240 (2007)

8 Gant, J.P., Gant, D.B.: Web portal functionality and state government e-services. In: The 35th Hawaii International Conference on System Science, pp. 1627–1636 (2002)

9 Kossak, F., Essmayr, W., Winiwarter, W.: Applicability of HCI research to e-government. In: 9th European Conference on Information Systems, pp. 957–968 (2001)

10 Dwivedi, Y.K., Williams, M.D.: Demographic influence on UK citizens' e-government adoption. International Journal of Electronic Government 5(3), 261–274 (2008)

11 Garcia, A.C.B., Maciel, C., Pinto, F.B.: A quality inspection method to evaluate E-government sites. In: Wimmer, M.A., Traunmüller, R., Grönlund, Å., Andersen, K.V. (eds.) EGOV 2005. LNCS, vol. 3591, pp. 198–209. Springer, Heidelberg (2005)

12 Beynon-Davies, P., Williams, M.D.: Evaluating electronic local government in the UK. Journal of Information Technology 18(2), 137–149 (2003)

13 Matera, M., Costabile, M.F., Garzotto, F., Paolini, P.: SUE inspection: an effective method for systematic usability evaluation of hypermedia. IEEE Transaction on System, Man, and Cybernetics – Part A: Systems and Human 32(1), 93–103 (2002)

14 Lee, S., Koubek, R. J.: The effects of usability and web design attributes on user preference for e-commerce web sites. Computers in Industry (2010); doi:10.1016/j.compind.2009.12.004

15 Verdegem, P., Verleye, G.: User-centered e-government in practice: a comprehensive model for measuring user satisfaction. Government Information Quarterly 26, 487–497 (2009)

16 Reddick, C.G.: The adoption of centralized customer service systems: a survey of local governments. Government Information Quarterly 26(1), 219–226 (2009)

17 Tolbert, C., Mossberger, K.: The effects of e-government on trust and confidence in government. In: Proceedings of the 2003 Annual National Conference on Digital Government Research, Digital Government Research Center, pp. 1–7 (2003)

18 Nielsen, J.: Heuristic evaluation: usability inspection methods, New York (1994)

19 Nielsen, J.: Deigning web usability: the practice of simplicity, Indiana, USA (2000)

20 Reddick, C.G.: Citizen interaction with e-government: from the streets to servers? Government Information Quarterly 22(1), 38–57 (2005)

21 Kappel, G., Pröll, B., Reich, S., Retschitzegger, W.: Web engineering, the discipline of systematic development of web applications. John Wiley & Sons, Chichester (2006)

22 Tuch, A.N., Bargas-Avila, J.A., Opwis, K., Wilhelm, F.H.: Visual complexity of websites: effects on users' experience, physiology, performance, and memory. International Journal of Human-Computer Studies 67(9), 703–715 (2009)

23 Brinck, T., Gergle, D., Wood, S.D.: Usability for the web: designing web site that work. Morgan Kaufmann Publishers, San Francisco (2002)

24 Tractinsky, N.: Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: Proceedings of ACM Conference on Human Factors in Computing Systems, pp. 115–122. ACM Press, New York (1997)

The Effects of Content Type and Presentation Style on User Experiences of Multimedia Content on a Tablet PC

Kari Kallinen[1], Jan Kallenbach[2], and Niklas Ravaja[1]

[1] Aalto University School of Economics, Center for Knowledge and Innovation Research
P.O. Box 2125, FIN-00076 Aalto, Finland
[2] Aalto University School of Science and Technology, Department of Media Technology
P.O. Box 15500, FIN-00076, Aalto, Finland
{Kari.Kallinen,Jan.Kallenbach,Niklas.Ravaja}@aalto.fi

Abstract. The present study examined how media type, presentation style, and user characteristics moderate people's media experiences while reading on and listening to news messages. We found that content type, presentation modality and user characteristics all moderated the responses to the news messages. We found for example that (1) entertainment news were experienced as more pleasant and activating, but factual news as more interesting, important, and trustworthy, (2) Audio news with text captions elicited better memory performance and higher presence than text or audio only news, (3) fast picture presentation elicited higher activation than slow pictures among younger users whereas slow picture presentation elicited higher activation than fast pictures among older users. The results demonstrate the complex mixture of user (such as age, and level of education), presentation (such as medium, modality, speed), and content (such as fact and entertainment) characteristics in media experiences.

Keywords: eReading, multimedia, user responses, emotion, memory performance, interest, presence.

1 Introduction

It has been suggested that electronic reading will become more widespread among media consumers and little by little paper-based newspapers and magazines will cease to exist. Regardless whether this is true or not: it is evident that reading from various electronic devices such as computers, laptops, mobile phones and electronic book readers has expanded very rapidly in the last few years. Publishers are facing new challenges as more and more of content is delivered to electronic devices.

In many aspects, electronic reading is in different from reading from a paper. The most notable differences are, on the one hand, the fact that readers can interact with the content/device and, on the other hand, that modern devices are becoming multifunctional with rich multimedia content that users are able to look, listen, modify and interact.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 466–475, 2011.

Previous studies have examined, for example, how people read text presented on a small display (Juola, Tiritoglu & Pleunis, 1995), how information visualization techniques can support reading of electronic documents (Hornbæk & Frøkjær, 2001), how memory for spatial location can be supported while reading from small displays (O'Hara, Sellen, & Bentley, 1999), and how background music and individual differences moderates the reading speed and experience while reading news from personal digital assistants (PDAs; see Kallinen, 2002; Kallinen, 2004).

However, it seems that most of the previous studies lack a broader theoretical framework for media experiences. In relation to the context of reading from computers (such as small displays), Kallenbach (2009) conceptualizes media experience on the basis of three qualities. First content quality refers to the extent of how well the message of mediated content can be conveyed to the reader or user. The second quality determining media experiences is the presentation quality. It refers to the extent of how contents are presented, displayed, or rendered. The third quality of electronic reading experiences is the quality of use. This refers to the usability of both the reading device itself as well as the interface displaying the reading contents. In addition to the content and medium related factors of media experience, media experience is strongly moderated by the context of the media (e.g., work, leisure, on move, at office etc.) use as well as the subjective factors of the users (e.g., preferences, attitudes, competence, age, profession etc.). Media experience can be seen as an interactive process of various components, such as the content, medium, content presentation and the individual user in a certain context at a certain time. In the present experiment we aimed to examine all these factors in an electronic reading setting. We were interested in the effects of content type (factual vs. entertainment), presentation type (text, audio, or text + audio with no-pictures, slow picture pace or fast picture pace) and individual characteristics of users (e.g., age and level of education) on the responses to news messages presented in a tablet simulation in a laptop. With regard to user responses, generally the factors cognition, emotion, and behavior are of importance when investigating subjective media experiences. We focused on these user responses in our study.

2 Research Problems and Expected Outcome

In the present study we were interested in the following research questions, among other things:

- How is different media content (entertainment vs. factual) experienced in an electronic reading situation?
- How does presentation type (text, audio, text + audio) moderate media experiences?
- What is the suitable presentation speed for pictures (fast vs. slow) for different media content?
- What are the individual differences in the media experiences?

We expected that content type, presentation mode and picture presentation pace all moderate media experiences. In regard to content type, given the nature of the news we expected that entertainment news would be eliciting more pleasantness and activation, whereas factual news would be considered as more important and trustworthy. In regard to the presentation style, given the connotation of daily media use, we expected that audio and pictures (especially fast pictures) would be more suitable for entertainment than factual news, whereas people may prefer factual content in connection with text. We also expected that the results would be moderated by the individual differences of the users (e.g., age and level of education), for example so that people with higher levels of education would have better memory performance than lower educated people in textual condition, given that they are likely more experienced in processing textual material.

3 Method

3.1 Participants

Thirty eight participants with varying fields of profession participated in the study in return for three movie ticket. Fourteen of them were men and twenty four were women; their age ranging from 19 to 46 years (M = 26.69). On average, participants had little experience with eReaders or other small hand held devices (excluding mobile phones), but they use computers, and read factual and entertainment news from computers quite frequently. All the subjects were fluent users of computers and had normal vision and hearing.

3.2 Stimuli

The stimuli material consisted of 18 Finnish entertainment (e.g., “Tässä ovat Diilin uudet yllätystuomarit”; ”Marja Tyrni myi kultaa” etc.) and 18 Finnish factual (e.g., ”EU-maat hyväksyivät Viron eurojäsenyyden”; ”Rankkasateet pahentavat edelleen Kiinan tulvia” etc.) news.

The news and the two content types were balanced so that the number of words and paragraphs per individual news item and per category was about equal. The mean word count was 131.00 for the 18 entertainment news and 132.12 for the factual news. On average there were 6.67 paragraphs in entertainment news and 6.82 in factual news. The mean audio length was 86.44 seconds for entertainment news and 91.1 seconds for factual news.

Each news item was presented in form of text, audio, or text with audio. In addition each item was paired with 3 relevant pictures whose pace of appearance on the display varies between a slow picture pace, a fast picture pace, and no picture. Examples from various stimuli presentations are illustrated in Figure 1 and Figure 2.

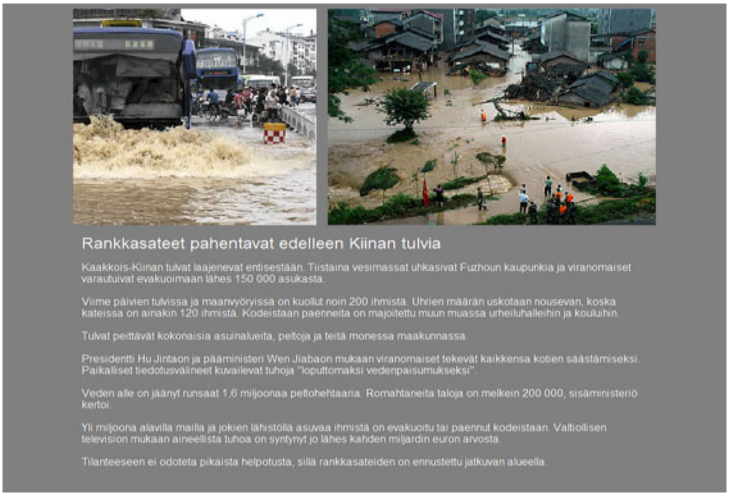


Fig. 1. A news item containing text with corresponding animated pictures

Fig. 2. A news item containing audio with captions and animated pictures

3.3 Measures

Memory Performance (Immediate and Delayed Recollection). Short term memory performance was measured by two multiple choice questions that were presented right after each news stimuli (immediate recollection). Longer term memory performance was measured by asking the participant list all news items they could remember (from any category) after the experiment (delayed recollection).

Self-Report Responses. In regard to media experience, the subject rated their subjective enjoyment and activation level, understandability, trustworthiness, interest, and importance of the news for themselves, as well as the overall satisfaction to the news "service" using 5-point Likert scales (from 1, e.g., unpleasant, uninteresting, unimportant etc. to 5, very pleasant, very interesting, very important etc.). In addition we collected various relevant background data of the participants, such as age, gender, level of education, previous experience with eReaders, and how used they are to reading factual and entertainment news from a computer.

Presence. To assess the feeling of presence we used a 7 item scale from the MEC-SPQ Questionnaire (Vorderer, Wirth, Saari, et al., 2004). The items (e.g., "I concentrated on the news"; "I did not notice the passing of time"; "I felt like I was a part of the environment in the news.") were measured by using a 5-point Likert scale ranging from 1 (I do not agree at all) to 5 (I fully agree).

3.4 Design and Analysis

The study design was Content (2 levels: entertainment, factual) x Presentation Modality (3 levels: text, audio, text and audio) x Picture Presentation (3 levels: no picture, slow pace, fast pace) within-subjects design. In each factors combination we had two news stories to minimize the effect of a particular story. The data for the two individual news stories representing the same category were pooled over and analyzed in SPSS using Liner Mixed Models (LMM).

4 Results and Discussion

The statistical analyses revealed over 40 significant main effects and 2-way interactions. The selected most interesting significant effects are described in the following chapters: main effects for content type, main effects for the content presentation style, interactions between content, modality, and picture presentation, and individual differences in responses.

4.1 Main Effects for Type of Content

The analyses showed main effects for content (factual vs. entertainment) in predicting users' immediate and delayed memory performance, pleasantness ratings, activation ratings, interest ratings, importance for oneself ratings, understandability ratings, and trustworthiness ratings (in all analyses $p<.001$).

The following Table 1 summarizes the results. The darker fill illustrates higher value and lighter fill lower value for the particular user response named in the most left column. Note that these values are relative between the two news types. For example, the lighter fill does not mean that factual news would be experienced as negative, but it merely illustrates that they are experienced as less pleasant than entertainment news.

As illustrated in the Table 1, entertainment content generated higher self-report pleasantness and activation than factual news. They were also considered as more understandable. Other results were perhaps less obvious: entertainment news elicited better immediate performance than factual news, but factual news were better remembered than entertainment news after a longer period of time. Also interestingly, even though entertainment news were considered as more pleasant, factual news were regarded as more interesting, important, and trustworthy. We think that these results reflect the nature of the news, and probable due this also the orientation of the participants on the news content (with a favor to factual news). Entertainment news had less important information value, they were perhaps more easily read/listened as they did not contain so much pure facts, and their content was more “harmless”. In contrast, factual news contained more often facts, and some of them were quite negative (e.g., nature catastrophe in china).

Table 1. Illustration of the effects of type of news on user experiences

User response	Content Type	
	Entertain. News	**Factual News**
MEMORY: Immediate recollection		
MEMORY: Delayed recollection		
PLEASANTNESS		
ACTIVATION		
INTEREST		
IMPORTANCE FOR ONESELF		
UNDERSTANDABILITY		
TRUSTWORTHINESS		

4.2 Main Effects for Content Presentation Style (Modality and Picture Presentation)

The analyses showed main effects for modality (text, audio, caption) in predicting users’ immediate memory performance (*p*=.048), activation ratings (*p*=.011), importance for oneself ratings (*p*=0.27), understandability ratings (p=.001), overall satisfaction ratings (*p*=.008), and presence ratings (*p*=.001). The following table illustrates the effects of modality on user experiences (see table 2). The darkest color fill is for the highest mean, second darkest color fill is for the second highest mean and lightest color fill is for the lowest mean. If two means are close to each other, they are illustrated with a same color fill.

As illustrated in the Table 2, the caption (audio + text) modality seems to elicit best short term memory performance, and highest level of activation, satisfaction, presence and positive emotion. Participants also evaluated news as most understandable and important when they were presented as audio with accompanying text. We think that many of these results reflect the task: participants were informed that they should study the material in order to be able to answer the questions about the news content. It was probably easiest to accomplish the task for news items with audio and captions because the information was given in two modalities: auditory and visual, as compared to audio or text only. In regard to presence (immersion into stimuli) it has been shown that usually sensory rich media tend to elicit higher presence than poor ones.

Table 2. Illustration of the effects of modality on user experiences

User response	Modality		
	Caption	Text	Audio
MEMORY: Immediate recollection			
ACTIVATION			
IMPORTANCE FOR ONESELF			
UNDERSTANDABILITY			
OVERALL SATISFACTION			
PRESENCE			

The analyses showed also main effects for picture presentation (no picture, slow picture pace, fast picture pace) in predicting users' delayed memory performance (p=.004), understandability ratings (p=.029), trustworthiness ratings (p=.038), and overall satisfaction ratings (p=.023). The following table illustrates the effects of modality on user experiences (see Table 3).

As illustrated in Table 3, participants rated the news with pictures as more trustworthy, understandable, and satisfactory. Again, the darkest color fill is for the highest mean, second darkest color fill is for the second highest mean and lightest color fill is for the lowest mean. If two means are close to each other, they are illustrated with a same color fill.

We think that, that the reason for this effect is that pictures illustrate the story, link the text/audio with easily remembered images, and create an imagery of the physical place where actions described in the news were taking place. Thus readers may have felt that news with pictures were more understandable. The difference between fast and slow pictures might be explained in terms of task effort: the fast pictures were "too fast", therefore disturbing the reading/listening whereas the slow pictures were ok (they illustrated the text/audio but did not take too much attention from reading/listening). Interestingly, even though participants considered news with pictures more understandable, their memory performance was worse in connection with news with pictures than news without pictures. It may be that as they were more difficult to understand, they were processed more thoroughly, which then showed in better memorization.

Table 3. Illustration of the effects of picture presentation on user experiences

User response	Picture presentation		
	No	Slow	Fast
MEMORY: Delayed recollection			
UNDERSTANDABILITY			
TRUSTWORTHINESS			
OVERALL SATISFACTION			

4.3 Interactions between Content Type, Modality, and Picture Presentation

Analyses revealed Content Type x Modality interaction in predicting interest (*p*=.045) and presence ratings (*p*=.031).

As illustrated in Figure 3 (left panel), text elicited highest interest ratings within factual news, but lowest for the entertainment news. As also illustrated in figure 3 (right panel), caption elicited highest presence ratings for both factual and entertainment news, but for text and audio the ratings were moderated by the content: for factual news text elicited higher presence than audio whereas for entertainment audio elicited higher presence than text. We believe that the reasons for these results are related to the connotations of traditional media and media use. Factual news are associated with a "serious" content and more "minimalistic" text-oriented presentation, whereas entertainment news may take more advantage of multimedia and audio forms of presentation.

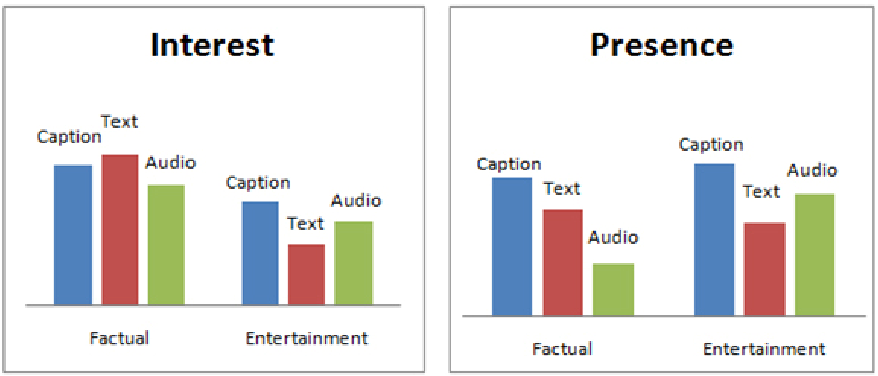


Fig. 3. Interest and presence ratings for factual and entertainment news as a function of presentation modality

4.4 Individual Differences in Responses

We found several individual differences in the responses, most importantly differences owing to participant's age and level of education. The following interactions were found in connection with the level of education: Content Type x Level of Education interaction in predicting interest (*p*=.032) and importance for oneself (*p*=.021) ratings; Modality x Level of Education in predicting immediate memory performance (*p*=.009), overall satisfaction (*p*=.001) and presence (*p*=.024) ratings. In sum, these interactions indicated the following:

– Factual news elicited more interest and importance than entertainment news especially among the subjects with higher levels of education.
– For higher educated participants caption elicited the best memory performance and highest satisfaction and presence ratings, text condition the second best memory performance and satisfaction and presence ratings, and audio the worst

memory performance and lowest satisfaction and presence ratings, whereas for lower educated the opposite was true: audio elicited better memory performance and higher satisfaction and presence ratings than text and caption.

The aforementioned results implicate that higher educated people are more factual and text oriented whereas in contrast lower educated may prefer entertainment content and auditory presentation.

In regard to individual differences related to age, we found a significant Content Type x Age interaction in predicting pleasant ratings (p=.019) and a Picture Presentation x Age interaction in predicting activation (p=.008) and understandability (p=.040). The results implicate the following:

- Especially young people rated the entertainment news as more pleasant than the factual news.
- For younger subjects fast picture presentation solicited higher activation ratings than slow pictures or no-pictures, whereas for older participants slow picture presentation elicited higher activation ratings than no-pictures or fast pictures.
- Similarly, for younger participants fast pictures elicited higher understandability ratings than slow pictures or no-pictures, whereas for older participants slow pictures elicited higher understandability ratings than fast pictures or no-pictures.

The aforementioned results may implicate a different habituation to news between the younger and the older subjects: young people are more used to and seek fast paced stimuli, whereas older people prefer news with slower picture pace.

5 Conclusions

In general, we found that both content type and presentation style exerted an influence on people's experiences of media stimuli in a situation of consuming eReading contents. We also found that these responses were moderated of individual differences owing to age and level of education. The results show that user responses to media stimuli are a complex mixture of user (such as age, personality, and reading habits), presentation (such as medium, modality, speed), and content (such as factual and entertainment news) characteristics. By examining the interaction of these different factors we can make predictions of user responses and adjust the content and presentation style to elicit more desirable media experience for example in terms of enjoyment, energy and interest. Thus, the result may also have practical value, given that it may make it possible to adjust the presentation of content in eReading devices to a particular user segment to optimize the media experiences. In addition, the results will improve our basic understanding of the different factors influencing on electronic reading related media experiences.

Acknowledgments. This study was supported by the Finnish Funding Agency for Technology and Innovation (TEKES) project "eReading".

References

1. Hornbæk, Frkjær, Hornbæk, K., Frøkjær, E.: Reading electronic documents: the usability of linear, Fisheye, and overview+detail interfaces. In: Proceedings of ACM Conference on Human Factors in Computing Systems, pp. 293–300. ACM Press, New York (2001)
2. Juola, J., Tiritoglu, A., Pleunis, J.: Reading text presented on a small display. Applied Ergonomics 26(3), 221–229 (1995)
3. Kallenbach, J.: Media Experience. In: Print & Hybrid Media, Finnish Paper Engineers' Association, Helsinki, Finland (2009)
4. Kallinen, K.: Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. Computers in Human Behavior 18(5), 537–551 (2002)
5. Kallinen, K.: The Effects of Background Music on using a Pocket Computer in a Cafeteria: Immersion, Emotional Responses, and Social Richness of Medium. In: Proceedings of CHI 2004 (2004)
6. O'Hara, K., Sellen, A., Bentley, R.: Supporting memory for spatial location while reading from small displays. In: CHI19'99 extended abstracts on Human factors in computing systems, Pittsburgh, Pennsylvania, May 15-20 (1999)
7. Vorderer, P., Wirth, W., Saari, T., Gouveia, F.R., Biocca, F., Jäncke, L., et al.: Development of the MEC spatial presence questionnaire (MEC-SPQ). Project report. Presence: MEC, IST-2001-37661 (2004)

Inherent Usability Problems in Interactive Voice Response Systems

Hee-Cheol Kim[1], Deyun Liu[2], and Ho-Won Kim[3]

[1] School of Computer Engineering/UHRC, Inje University, Korea
[2] School of Computer Engineering, Inje University, Korea
[3] School of Management, Inje University, Korea
heeki@inje.ac.kr, ludiyn@gmail.com, nannaya829@naver.com

Abstract. We are often required to use interactive voice response (IVR) systems during daily life. However, users still find these systems difficult to navigate. By identifying the problems with IVR systems, we can work to make them more convenient and easier to use. Based on the results of a literature review and an experimental study, this paper identified four major problems inherent to the use of IVR systems: linearity, transience, ambiguity, and minimal feedback. These issues cannot be completely avoided within IVR systems, but by understanding how the problems arise, designers can try to minimize them.

Keywords: Human–computer interaction, interactive voice response systems, usability.

1 Introduction

Two kinds of interactive voice response (IVR) systems are commonly used: telephone data entry (TDE) and automated speech recognition (ASR). Of the two, TDE is more widely used for checking airline reservations, telephone banking, surveys, automatic online ordering, etc. Users can access required voice information using a touch-tone interface (TTI; see Fig. 1). However, these systems involve many usability problems. For example, it may be difficult for a user to find the required information even using the entire menu and all system prompts.

Fig. 1. Touch-tone interface with 12 standard keys

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 476–483, 2011.

It is important to identifying fundamental problems related to interaction with IVR; if IVR designers are aware of problems, they can work to develop better and more usable systems. We conducted a literature review and an experimental explorative study in which 10 participants used IVR systems to perform two different tasks (checking a monthly bill and cancelling an insurance contract). Subjects also completed a questionnaire and interviews about the usability of IVR systems.

Section 2 presents background information about computer telecommunication integration (CTI) and IVR. Section 3 describes our experimental study and previous research about the usability of IVR systems. Section 4 analyzes four problems inherent to IVR system usability: linearity, transience, ambiguity, and minimal feedback, and finally we present our conclusions.

2 Background

2.1 IVR Systems

With the development of information and communication technology, CTI technology has linked computer and telephone technologies and is widely used in call-center service systems. CTI systems can include auto-answer, voice mail, fax, short-message service (SMS), automatic switching, sound recording services, etc. They have become an indispensable business tool in the competitive environment of modern business.

IVR systems are an important part of CTI; their main function is to guide users using phone menus and prompts, automatic response, automatic number switching, data search, e-mail, fax, and other functions, as shown in Fig. 2.

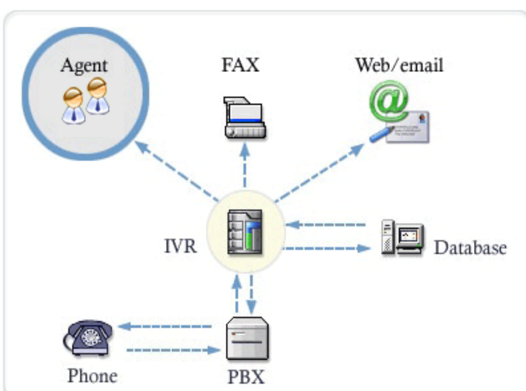


Fig. 2. Functions of IVR systems

Two main types of IVR systems are currently used: TDE and ASR. These vary in terms of how users input responses [1]. ASR systems do not yet have the ability to accurately recognize voices, so their use is limited; TDE systems are much more common. This paper focuses on TDE systems, which are most frequently used for checking orders and flight status, telephone banking, surveys, automatic online ordering, etc. IVR systems can provide required voice information for users. Figure 3 displays the workflow from initialization to disconnection.

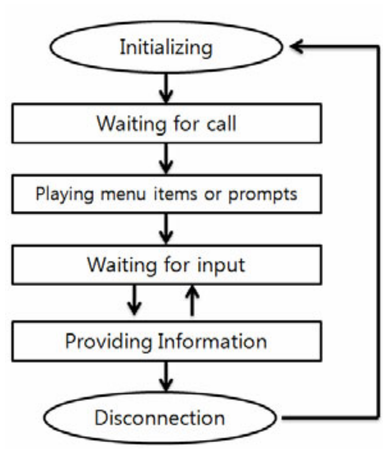


Fig. 3. Workflow of IVR system

3 Literature Review and Study

Our literature review and experiment revealed that IVR systems have many problems that inconvenience users. The body of literature about the usability of IVR systems provides numerous solutions for overcoming the various deficiencies of IVR systems. Some studies have found that the depth of a menu should not exceed four levels, and the menu width should not exceed nine elements. Gardner-Bonneau suggested balancing the depth and width of a menu by designing a graph structure [2]. Users are more likely to get lost in an IVR system if paths are ambiguous, so these should be avoided [3][4]. One study found that including an additional component increases the controllability of IVR systems [5], and skipping and scanning keys can be also used to reduce navigation time [6]. Some systems allow users to set personal menus and to access these directly in the hierarchy without going through the various layers [7][8]. Insertion of speech-based earcons or non-speech earcons reportedly improves navigation through auditory menus [9][10][11]. These methods can help improve the usability of IVR systems, but more research is needed to assess the problems within IVR systems and possible solutions. To date, few usability-related studies have been conducted, and the level of user satisfaction with IVR systems is low.

For our study, we asked 10 participants to use the IVR systems of one mobile company and one insurance company in Korea to check charges on the previous month's telephone bill and to buy and cancel an insurance policy. After completing the tasks, participants completed a questionnaire and an interview about how they liked each system in terms of usefulness and usability. The goal was to identify ways to improve the usability of IVR systems. The experimental data revealed usability problems related to IVR systems and suggested ways to overcome them. For example, some participants noted that they found it difficult to find the menu they wanted. They found it difficult to remember the entire menu over time and were forced to have it repeated, but they were not able to skip unnecessary content. Sometimes they were also unable to even move back to a higher-level menu. Some participants suggested using a shortcut button to improve the control speed. They also suggested making the menu more concise.

4 Four Inherent Problems of IVR Systems

Based on the literature review and our experimental data, we identified four problems inherent to IVR systems:

4.1 Transience

There are four basic kinds of human memory: sensory memory, short-term memory, working memory, and long-term memory [12]. Sensory memory is used to create an overall sensory picture, which is usually stored for 2–4 seconds. Short-term memory is used as temporary storage for sensory (including auditory) input. It lasts slightly longer than sensory memory. Short-term memory is used to combine sensory inputs and to focus them for a short time. It dissipates quite rapidly, lasting only a few seconds. Working memory provides a somewhat larger perspective than short-term memory, although in some cases, the two are considered to be identical. Long-term memory involves relatively permanent memory traces that are stored beyond short-term memory.

Listening is controlled by short-term memory [13]. For example, by the time we hear "Press 1 for English services," the instruction has ended. This means that we use short-term memory to memorize items and complete tasks in an IVR system. So, IVR systems require that we use memory to complete the tasks. However, our brains will not store this information for long, and in time, it will be forgotten.

In many cases, users must listen to the instructions again because they immediately forget what they heard. One participant said that the menu prompts were very long, so he understood them clearly but could not remember them easily. He was not able to find the task menu that he needed. This shows that it is difficult to receive and remember a large amount of data from an IVR system at one time. Transmitting information via voice is consecutive, but human memory is transitory.

Because of the transience of memory, users find it difficult to remember system menus and prompts that are too long. They also find it difficult to make choices and deal with tasks in an IVR system, leading to low satisfaction levels and usability in systems that use voice-only interfaces. In this respect, therefore, transience can be considered an important inherent problem of IVR systems. In making an IVR system understandable, easy, and controllable, it is important to minimize problems caused by transience.

4.2 Linearity

We can use at least two different approaches to convey to someone (A) the message: “Press 1 for English services. Press 2 for Chinese services. Press 3 for Spanish services." We could present the information in written form (Figure 4), or we could present the information verbally (Figure 5). The first way involves a graphical user interface (GUI), and the second involves a voice user interface (VUI), which is considered a linear method.

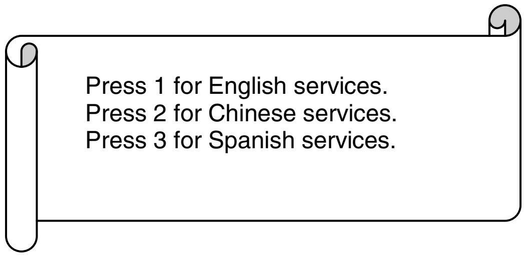


Fig. 4. Graphical User Interface

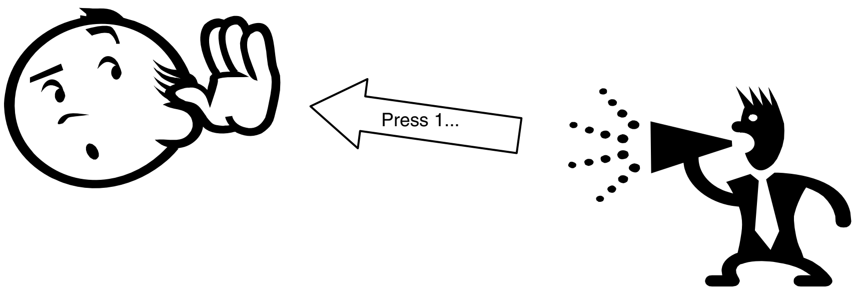


Fig. 5. Voice User Interface

A linear type of interaction means that users can only navigate information in a given order [15]. Speech and audio interfaces must be sequential [16], whereas visual interfaces can be simultaneous. For example, a user who wants to perform the task in Spanish must first spend time listening to "Press 1 for English services" and "Press 2 for Chinese services" (Fig. 5) before finally selecting the Spanish services. This wastes time, unlike a GUI mode that can display all messages at one time (Fig. 4).

One participant in our experiment said, "After the telephone was connected, I heard an advertisement first. I was very surprised, and spent a long time listening to the advertisement. I was very angry at that time." Another participant said, "The purpose of my call is to solve the problem quickly, but I must listen to the menus from the first one to the end, and make a choice after finding what I need." The ear cannot browse around a set of recordings, unlike the eye, which can scan a screen of text and images at a glance. Therefore, it is important to consider the linearity problem when designing systems based on speech and audio interfaces.

4.3 Ambiguity

Speech can be ambiguous during communication; this is related to another problem inherent to IVR. During conversation, we often say, "I beg your pardon, could you repeat that?” because we did not hear or understand something. Many letters (such as B and D) sound very similar [17], and sometimes two words that are pronounced the same have spellings and meanings that differ greatly (e.g., "4" and "for"). Although we can usually judge the meaning of a word within the context of a sentence, we may not understand if only one word is spoken.

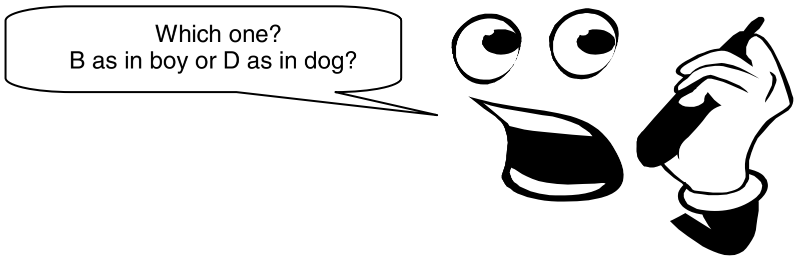


Fig. 6. Ambiguity of speech

The ambiguity of speech is exemplified frequently in daily life. When we listen to a new song without seeing the lyrics, it can be very hard to understand some words. When we watch DVDs without captions, it can sometimes be difficult to follow the dialogue in the movie.

To resolve the ambiguity of speech, many DVDs now include captions at the bottom of the screen, and many MP3 players display lyrics as a song plays. One participant in the experiment was angry about the insurance company’s IVR system, saying, “Because the speech is intermittent and the pronunciation of the words is ambiguous, I couldn’t hear the menus and prompts well. And then I was in a bad mood and complained to the service staff directly.” This example shows that it is important to address ambiguity in IVR systems.

4.4 Minimal Feedback

Feedback helps a user know the working state of a system. It is extremely difficult for a system with only a voice interface to convey content and information about the system’s working state simultaneously. The lack of visual feedback in IVR systems can make users feel less in control [18], as users have only poor real-time auditory feedback while they are completing a task.

In contrast, users of GUI systems can deal with tasks quickly and feel confident because they receive abundant real-time feedback. Figure 7 shows how a GUI system can show users its working state including its rate of completion, elapsed and remaining time, compression rate, and processing time. This allows the user to understand the working state within seconds.

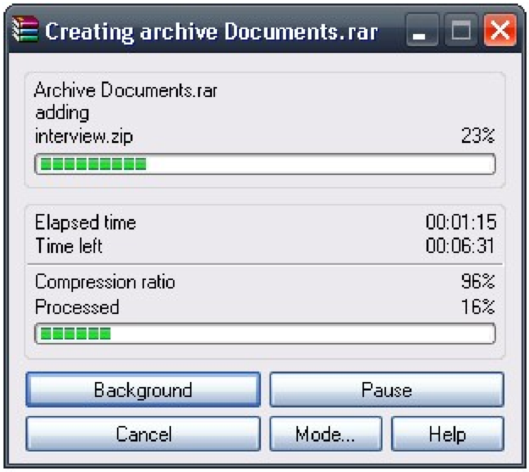


Fig. 7. Compression process in WinRAR

This example shows how a GUI system can convey many kinds of information at once, whereas a VUI system provides only one kind of information at a time. However, feedback time must also be short because of the limited capability of our working memory [19]. Minimal feedback is a fundamental problem of VUI systems that prevents users from receiving adequate feedback. This problem is also closely related to the slow speed of speech output [20].

The minimal level of feedback forces users to spend more time to complete a task. One participant said, "I like to use the web browser to do the tasks because I can get the exact information when I am doing my tasks. But it is difficult to get much information at the same time when I am connected the IVR system." The fact that users were unable to obtain much useful feedback is a very serious defect of IVR systems and should be addressed when designing IVR systems.

5 Conclusions

IVR systems can provide users with a variety of voice services. IVR systems are widely used because companies can reduce costs by employing IVR systems for customer care, and phones are now accessible to almost every user. However, few studies have focused on the usability of IVR systems. By emphasizing aspects of usability, this study identified inherent usability problems in IVR systems, including linearity, transience, ambiguity, and inadequate feedback. These findings can be used to help designers develop better and more user-friendly IVR systems.

Acknowledgments. This work was supported by the Korea Research Foundation (KRF) grant funded by the Korea government (MEST) (No: 2009-0076772).

References

1 Groves, C.: ABS Forms Design Standards Manual Forms Design Standards Manual– IVR Design Standards, ch. 30 (Novomber 2004)
2 Gardner-Bonneau, D.: Human factors and voice interactive systems. Kluwer Academic, Norwell (1999)
3 Balentine, B.: Re-Engineering the speech menu. In: Gardner-Bonneau, D. (ed.) Human factors and voice interactive systems, pp. 205–235. Kluwer Academic, Norwell (1999)
4 Balentine, B., Morgan: How to build a speech recognition application. Enterprise Integration Group Inc., San Ramon (1999)
5 Karat, C.M., Halverson, M., Horn, D., Karat, J.: Patterns of entry and correction in large vocabulary continuous speech recognition systems. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 568–575. ACM Press, New York
6 Resnick, P., Virzi, R.A.: Kip and scan: Cleaning up telephone interfaces. In: Proceedings of ACM CHI 1992, pp. 419–426 (1992)
7 Peer, M., Shajahan, P., Irani, P.: Improving Navigation in Touch-Tone Interfaces. In: International Symposium on Human Factors in Telecommunication, 20031201-20031204, Berlin, DE
8 Irani, P., Shajahan, P., Kemke, C.: VoiceMarks: restructuring hierarchical voice menus for improving navigation. Int. J. Speech Technol. 9, 75–94 (2006)
9 Walker, B.N., Nance, A., Lindsay, J.: Spearcons: speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the 12th International Conference on Auditory Display (ICAD 2006), pp. 63–68 (2006)
10 Brewster, S.A.: Navigating telephone-based interfaces with earcons. In: 12th British Computer Society HCI Conference, Bristol, England, August 12-15 (1997)
11 Shajahan, P., Irani, P.: Representing Hierarchies Using Multiple Synthetic Voices, pp. 885–891. IEEE Computer Society, Los Alamitos (2004)
12 Forester, L.: Implications of Research on Human Memory for CALL Design. CALICO Journal 20(1) (September 2002)
13 Schnelle, D., Lyardet, F.: Telecooperation Group, and Darmstadt University of Technology Hochschulstraÿe 10, Voice User Interface Design Patterns
14 Dey, A.K., Catledge, L.D., Abowd, G.D., Potts, C.: Developing Voice-only Applications in the Absence of Speech Recognition Technology. Georgia Institute of Technology (April 2003)
15 Kim, H.-C.: Weaknesses of Voice Interaction, Networked Computing and Advanced Information Management. In: NCMapos 2008: Fourth International Conference, 2nd edn., pp. 740–745 (2008)
16 Arons, B.: Hyperspeech: Navigating in Speech-Only Hypermedia. In: HYPERTEXT 1991: Proceedings of the Third Annual ACM Conference on Hypertext, pp. 133–146 (1991)
17 Yankelovich, N., Levow, G.A., Marx, M.: Designing speechActs: Issues in Speech User Interface. In: CHI 1995 Conference on Human Factors in Computing Systems, May 7-11, Denver, CO (1995)
18 Stifelman, L.J., Arons, B., Schmandt, C., Hulteen, E.A.: VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker, pp. 179–186. ACM, New York (1993)
19 Cohen, M.H., Giangola, J.P., Balogh, J.: Voice User Interface Design, February 6. Addison Wesley, Reading (2004)
20 Yankelovich, N., Levow, G.A., Marx, M.: Designing speechActs: Issues in Speech User Interface. In: CHI 1995 Conference on Human Factors in Computing Systems, Denver, CO, May 7-11 (1995); Dey, A.K., Catledge, L.D., Abowd, G.D., Potts, C.: Developing Voice-only Applications in the Absence of Speech Recognition Technology, Georgia Institute of Technology (April 2003)

Effect of Aesthetic Design Elements on Tabletop Display Interaction

Hyunglae Lee[1], Hyunjin Shin[2], and Ji-Hyung Park[3]

[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, USA
[2] Convergence Biz Division, LG Uplus Corp., South Korea
[3] Intelligence and Interaction Research Center, Korea Institute of Science and Technology, South Korea
hyunglae@mit.edu, zeze0@lguplus.co.kr, jhpark@kist.re.kr

Abstract. Recent technology developments in the field of tabletop display systems have provided new types of user experiences that were unavailable in traditional vertical displays. While substantial amount of work has been done on new technologies and usability tests for a tabletop display system, not much attention was focused on aesthetic design aspect. In this paper, we studied the effect of aesthetic design elements on tabletop display interactions, especially focusing on users' emotional satisfaction and usability. For this purpose, two different tabletop display systems, one aesthetically well designed with 5 selected design elements (color, typography, form, graphic, and sound) and the other not, were tested in an intelligent collaborative environment. The user study showed that an aesthetically well-designed tabletop system not only influenced users' emotional satisfaction but also usability by enhancing the abilities in human perception, and eventually affected overall user experience with the system.

Keywords: Aesthetic design elements, Tabletop display.

1 Introduction

In recent years, tabletop displays, often called as interactive tabletops and surfaces, have been getting a great deal of attention. Display technologies, such as projectors, flat panel displays (LCD, LED, and so on), coupled with sensing technologies, enable direct and natural human-computer interaction with hands as an input device. While much of the research attention has focused on developing new technologies capable of new types of interaction, for example, multi-touch interaction [1], new interfaces integrated with existing interfaces [2–4], little work has been done on the aesthetic aspects of tabletop display design [5], which may significantly influence user experiences. On the other hand, researches on other types of system have supported the importance of aesthetic design elements in human-computer interaction [6–8]. Experiments with different types of ATM layout revealed that the apparent usability is strongly affected by the aesthetic aspects rather than the inherent usability [6, 7].

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 484–489, 2011.

Another research on Web design showed that simplicity and beauty affect the user experience and interpretation of the design [8].

In this work, we studied the effect of aesthetic design elements on tabletop display interaction, especially focusing on users' emotional satisfaction and usability. At this stage of research we only investigate the overall effect of aesthetically oriented design resulted from the combination of several aesthetic design elements, not the effect of individual element on user experiences.

2 Tabletop Display Environment

A collaborative working environment called "Intelligent and Responsive Space (IRS) [9, 10], which consists of a tabletop display, three surrounding wall screens and mobile devices, was used as a test environment in this study. This space is utilized mainly for a group meeting or presentation. The tabletop display system, which plays a key role in this environment, was built horizontally by using a large display (55 inch LCD unit) and an infrared LED touch panel. Users can operate this system with bare fingers by manipulating a virtual keyboard/mouse and handling interactive menus and icons.

Two versions of interactive tabletop display systems were used in this study (Fig.1). The first version was designed mainly focusing on technology development [9]. The second version incorporates aesthetic design elements [10]. The 2nd version uses the same technology and has the same functional aspect, and the only difference is the design aspect.

Fig. 1. Two versions of tabletop display systems (top: 1st version, bottom: 2nd version with consideration of aesthetic design elements)

3 Aesthetic Design Elements

To make aesthetically well-designed system, aesthetic design elements for the tabletop display interaction environment were extracted based on a research about human cognitive response to an artifact [11]. According to his research, aesthetic responses in human cognitive process are adaptation for detecting physical features such as color, tone, shape, pattern, etc. These aesthetic responses through sensory interface are immediate and involuntary, apparently distinguished from analytical responses which usually take longer. These responses may affect the further determination of preferences, thus a positive aesthetic response is more likely to lead to a positive ultimate preference than negative. We selected target design elements that may affect aesthetic response in tabletop display interaction. Five elements were selected: Color, typography, form, graphic, and sound.

Color plays one of the most important roles in recognizing an artifact. Proper selection of color scheme and well designed combination of them attract users' attention [12], improve readability [13] and increase participant [14]. In an interface design, however, color should not be overused which usually causes heavy cognition load. We chose a minimal number of color sets (4 for background and 4 for foreground such as icons and commands) to emphasize the distinction between foreground and background and keep the unity of design (Fig.2).

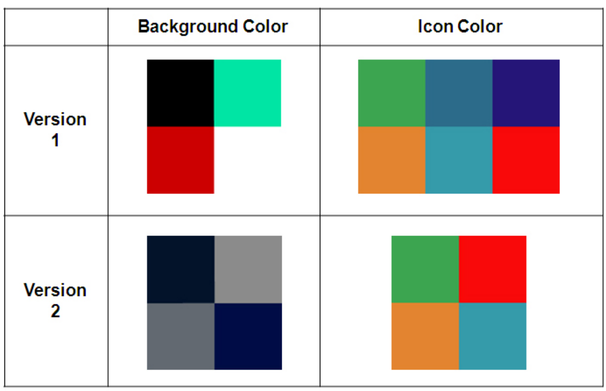


Fig. 2. Two different color sets used in the 1st version (left) and the 2nd version (right). In the 2nd version, background colors were chosen to have similar color tone to prevent distraction.

Since a tabletop display is a large horizontal display unlike traditional vertical displays, some displayed objects or texts proximal to one person may be distal to others or even reversed. Thus, readability and especially legibility can be significantly influenced by the selection of typography. According to the guideline suggested in [15], we selected regular type san serif fonts (Fig.3).

We designed the form of tabletop exterior according to the Gestalt laws that humans tend to order our experience in a manner that is regular, orderly, and simple. Compared to the 1st version, all of the peripheral components were enclosed in the exterior box. This design draws user attention toward the main display and eliminates distractions by surrounding structures or peripheral devices (Fig.1 bottom left).

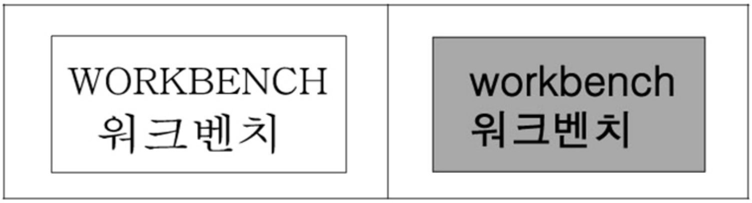


Fig. 3. Two different typography sets for the 1st version (left: Serif font) and the 2nd version (right: Sans-serif font)

Considering that the tabletop display is usually used by multiple users, and both input and output occur on the same display, proper graphic representation for information visualization and manipulation are required. We used interactive animations and special effects helping natural interactions, in the limit not to cause confusion in recognition and evoke cognitive fatigue. As a final design element, proper sound feedback and sound effects accompanied by animations were integrated to improve user responsiveness and give emotional satisfaction.

4 User Study and Results

We performed user studies using these two versions of tabletop to examine the effect of aesthetic design elements on using tabletop display systems. As mentioned above, there is no difference between the two versions in the functional aspect, but only in the design aspect.

We recruited 24 subjects (13 male and 11 female; age 24 to 38) having no experience of using tabletop display systems for the user study. First, users were asked to evaluate the aesthetic satisfaction on two systems after watching the experimenter's demos of the two versions about 30 minutes. In all evaluations, the 7-point Likert scale (1=strongly dislike, 4=undecided, 7=strongly like) was used. In the experiment, each subject was asked to perform 8 simple tasks. We designed the first half of the tasks to assess user cognitive load in performing tasks, and the second half to evaluate manipulability. Completion time for each task was checked throughout experiments. Before running the experiment, each user was trained for about an hour how to use two systems. After completing the tasks, the subject answered the evaluation questions asking about emotional satisfaction and overall satisfaction on each version.

In data analysis, we performed paired t-Tests on each evaluation item to compare two versions (Table.1). Both in aesthetic and emotional satisfaction evaluations, users rated the second version higher than the first version ($p<<0.05$). In addition, we found that there exists high correlation between aesthetic evaluation and emotional satisfaction (Pearson correlation coefficient R^2=0.61 and 0.71 for version 1 and 2, respectively). In overall satisfaction on the system, users also gave higher points to the second version over the first version ($p<<0.05$).

Table 1. Evaluation results: Mean of 24 subjects (Within parentheses: Standard deviation)

Evaluation Item	Version 1	Version 2
Aesthetic Satisfaction	4.79 (0.44)	5.82(0.52)
Emotional Satisfaction	4.48 (0.87)	5.61(0.40)
Overall Satisfaction	4.42 (0.49)	6.03 (0.45)

In particular, for the second version, evaluation for overall satisfaction was higher ($p<0.05$) than the evaluation for emotional satisfaction, while there was no statistical difference ($p>0.05$) for the first version. This implies that aesthetic design elements may affect not only emotional satisfaction but also usability. The result on task completion time comparison (Fig.4) strongly supports this argument. In all 8 tasks, task completion time in the second version was lower than that of the first one. In addition, time reduction in the first 4 tasks was higher than the last 4, which means aesthetic design elements greatly affect cognitive load reduction rather than improvements of manipulability.

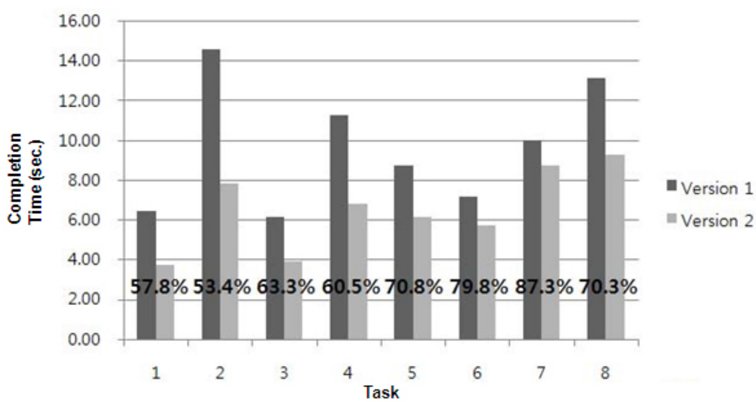


Fig. 4. Task completion time comparison for each task

5 Discussion

In this paper, we identified that aesthetic design elements play a very important role in human-computer interaction in a tabletop display environment. The user study showed that an aesthetically well-designed tabletop system not only influences users' emotional satisfaction but also usability, and eventually affects overall user experience with the system. Before running experiments, we expected that the emotional evaluation for the aesthetically well designed system (2nd system) would be higher than the original one (1st system). However, contrary to our expectation, the result showed that there is no statistical difference between aesthetic and emotional evaluation after experiencing the system. We thought this was due to the fact that even the 1st system scored high value (4.79) in aesthetic evaluation. We may evaluate the effect of aesthetic elements on the emotional satisfaction more clearly by testing two different systems with more distinction in aesthetic aspect. As future directions, we need to design and perform extensive user studies to identify the impact of each aesthetic design element on optimal user experience.

References

1 Han, J.: Low cost multi-touch sensing through frustrated total internal reflection. In: Proc. 18th Annual ACM Symposium on User Interface Software and Technology (UIST 2005), Seattle, pp. 115–118 (2005)
2 Lee, H., Jeong, H., Lee, J., Yeom, K., Shin, H., Park, J.: Select-and-Point: a novel interface for multi-device connection and control based on simple hand gestures. In: Proc. CHI 2008, Florence, pp. 3357–3362 (2008)
3 Everitt., K., Shen., C., Ryall, K., Forlines, C.: MultiSpace: Enabling Electronic Document Micro-mobility in Table-Centric, Multi-Device Environments. In: 1st Annual IEEE International Workshop on Horizontal Interactive Human-Computer System (Tabletop 2006), Adelaide, pp. 27–34 (2006)
4 Pattern., J., Ishii, H., Hines, J., Pangaro, G.: Sensetable: a wireless object tracking platform for tangible user interfaces. In: Proc. CHI 2001, Seattle, pp. 253–260 (2001)
5 Schiphorst, T., Motamedi, N., Jaffe, N.: Applying an Aesthetic Framework of Touch for Table-top Interactions. In: 2nd Annual IEEE International Workshop on Horizontal Interactive Human-Computer System (Tabletop 2007), Rhode Island, pp. 71–74 (2007)
6 Kurosu, M., Kashimura, K.: Apparent usability vs. Inherent usability. In: Proc. CHI 1995, Colorado, pp. 292–293 (1995)
7 Tractinsky, N.: Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In: Proc. CHI 1997, Atlanta, pp. 115–122 (1997)
8 Karvonen, K.: The Beauty of Simplicity. In: Proc. Conference on Universal Usability (CUU 2000), Arlington, pp. 85–90 (2000)
9 Park., J., Yeom., K., Ha., S., Park., M., Kim., L.: An Overview of Intelligent Responsive Space Technology in Tangible Space Initiative. In: Proc. Pervasive 2006, Dublin, pp. 523–232 (2006)
10 Lee, H., Jeong, H., Lee, J., Yeom, K.-W., Park, J.-H.: Gesture-based interface for connection and control of multi-device in a tabletop display environment. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 216–225. Springer, Heidelberg (2009)
11 Ulrich, K.T.: Design: Creation of Artifacts in Society, ch. 5 (2010)
12 Turatto, M., Galfano, G.: Attention capture by color without any relevant attentional set. Perception and Psychophysics 63(2), 286–297 (2001)
13 Garcia, M.L., Caldera, C.I.: The effect of color and typeface on the readability of on-line text. Computers and Industrial Engineering 31(1), 519–524 (1996)
14 White, J.V.: Color for Impact. Strathmoor Press (1997)
15 Michael, B., Liao, C.H., Melissa, M.: The effects of Font type and Size on the legibility and reading time of online text by older adults. In: CHI 2001, Seattle, pp. 175–176 (2001)

Effects of Presence on Causing Cybersickness in the Elderly within a 3D Virtual Store

Cheng-Li Liu[1] and Shiaw-Tsyr Uang[2]

[1] Department of Management and Information Technology
Vanung University, Taoyuan, Taiwan
johnny@vnu.edu.tw
[2] Department of Industrial Engineering and Management
Minghsin University of Science & Technology, Hsinchu, Taiwan

Abstract. Along with elderly society's coming, computers with an internet connection used at home can provide this population a new channel to access to information and services, and can also be used to manage internet shopping tasks. One of the primary advantages of virtual environment (VE) technology applied in web shops is its ability to provide a 3D perspective to customers for more real sense on goods and shopping environment. The presence is a great appeal for the elderly experienced in the 3D virtual store. How to improve the presence of 3D virtual store is a challenge for the shopkeepers and programmers. But, whether it is easy to produce cybersickness in the elderly when the presence is improved, or not? The objectives of this study are to explore the effects of depth perception cues and display modes on presence and incidence of cybersickness in the elderly within a 3D virtual store. The results show that participants in high level depth perception cues with 3D monitor (with circular polarized glasses) will experience greater presence than other displays (HMD and TFT-LCD). However, a feeling of presence may add to the incidence of cybersickness($r = 0.671$, $p = 0.000$). Although the feeling of presence in TFT-LCD display is poor than 3D monitor and HMD, the symptoms of cybersickness are the slightest.

Keywords: 3D virtual stores, Elderly, Presence, Cybersickness, Depth perception cues, Stereopicture.

1 Introduction

1.1 General Introduction

With rapid development of internet, online shopping has become a way of locating oneself within today's culture based on what people purchase and how they use their purchases; indeed, it has become quite popular since its emergence on the internet. Along with elderly society's coming, many older adults (i.e., the "silver tsunami" generation) have problems performing daily tasks because of restricted mobility, lack of transportation, inconvenience, and fear of crime [4]. Computers with an internet connection used at home can provide this population a new channel to access to

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 490–499, 2011.

information and services, and can also be used to manage internet shopping tasks. However, the traditional web shops introduce commodities only by 2D pictures and descriptive catalogue, which fall short in terms of reality and the interaction with goods. This kind design with poor reality and interaction will influence customers' real shopping experiences; and what is more, they minimize customers' desire to shop. Therefore, the elderly may find the barriers so great to prevent effective communication and shopping taking place [9]. Nowadays, such problems can be solved utilizing the technology of virtual environments (VEs). One of the primary advantages of VEs technology applied in web shops is its ability to provide a 3D perspective to customers for more real sense on goods and shopping environment. Therefore, we are deeply convinced that web 3D virtual store will be more and more popular in the future, and the older population would be growing rapidly worldwide and becoming an increasingly important demography in online shopping [10].

1.2 Presence and Depth Perception Cues of 3D Virtual Store

As we know the 3D virtual store is different from the common web store. The lifelike feeling of goods may hold a special attraction for the group of elderly. Therefore, it is a most important task for VE designers to develop an illusion of being "present" in a VE [24]. Several researchers have found that presence is generally regarded as a vital component of VEs for users to experience and interact with good sense in real time [17, 19]. Presence has been identified as being the defining characteristic, a design ideal or a desirable outcome of VE participation [23, 26]. Witmer (1998) defined presence as the subjective experience of being in one place or environment, even when one is physically situated in another [27]. Freeman et al. (1999) also described presence as the observer's subjective sensation of being there in a remote environment [5]. Lombard and Ditton (1997) proposed that presence is as a perceptual illusion of non-mediation involves continuous responses of the human sensory, cognitive and affective processing systems [15]. Therefore, a good presence of 3D virtual store would make users to immerse and involve in the shopping situation. The presence improving will be a challenge for shopkeepers and programmers [16].

There are some factors influence the degrees of presence within a VE. Perception of self-inclusion is a main factor [18]. Wickens et al. proposed that people would use a variety of depth perception cues to feel the shape and distance of objects within 3D environment [25]. Depth perception cues can be perceived via monocular depth cues, binocular depth cues and oculomotor cues. Monocular depth cues provide an equivalent percept to both eyes; the cues are equally effective whether using one or both eyes. Binocular cues, on the other hand, take advantage of both eyes by allowing each eye to receive slightly offset views of the same visual scene. Oculomotor cues occur via accommodation and convergence to involve combining visual and proprioceptive information from the eye to derive information related to distance.

Several researchers reported that monocular cues are the most important cues to depth in simulator displays [1, 6]. However, in the case of the near-view 3D display condition, the effect of the monocular accommodation increases with the blurring effect, so viewers can see unclear 3D images. Therefore, this is an inherent problem with good 3D image quality at the near-view. Additionally, binocular disparity is a binocular cue which creates the phenomenon of stereopsis and affected by the stereoscopic properties of a display. Hale and Stanney found that a lack of binocular disparity may reduce one's perception of presence within a VE compared to that of normal sighted individuals. A small degree of binocular disparity creates stereopsis, a compelling impression of depth, as the two retinal images are fused together to form one single image [3]. Because stereopsis provides an enhanced perception of depth, stereoscopic displays (e.g. head mounted display (HMD), non-immersive (high resolution, stereographic ready monitors with circular polarized glasses) and auto-stereoscopic displays (defined as a 3D visual display that does not require viewing aids)) are thought to contribute to increased presence [8].

1.3 Cybersickness and Depth Perception Cues

Cybersickness is an illness for some users to exhibit symptoms that parallel symptoms of classical motion sickness both during and after the VE experience, in that it is most probably brought about by a sensory conflict between the three major spatial senses: the visual system, vestibular system, and (non-vestibular) proprioception [14, 20]. The main symptoms of cybersickness are eye strain, disorientation and nausea [13, 22]. Several researchers have found that 80% to 95% of users will experience some level of disturbance during exposure to a VE, with between 5% and 30% experiencing symptoms severe enough to discontinue exposure [21].

Some researchers found that there are some problems which have been associated with visual displays inducing cybersickness. For example, Howarth (1996) found that exposure to stereoscopic displays will increase symptoms of cybersickness (e.g. eye strain, blurred vision), the reason is oculomotor disturbance due to the mismatched oculomotor cues by such displays [7]. In addition individuals with low-stereovision may be unable to completely fuse these images together (i.e. presented with two different visual scenes), and lead to increased oculomotor disturbances [6]. These studies showed that the contributing factors of cybersickness are not only sensory conflict or postural instability but also depth perception cues.

1.4 Objectives

The technology used to create virtual environment has come a long way in providing immersive VE experiences. Through a combination of both hardware and software features, computer graphics systems allow the possibility of creating 3D virtual environments which can appear to be quite realistic and, in addition, can improve specific spatial task performance with good 3D image quality. When depth perception cues are improved with good 3D image quality and stereoscopic displays, however, whether the cybersickness for the elderly during exposure within a 3D virtual store stereoscopic displays will be easily caused or not? Therefore, the purpose of this study was to 1) understand clearly the effect of depth perception cues and display modes on presence and cybersickness in the elderly within a virtual store, 2) discuss the influence of presence that contribute to cybersickness.

2 Method

2.1 Participants

There were 60 people (average age of 65.3 years) selected to participate in the experiment. They were paid a nominal NTD500 as compensation for their time. All participants were fully informed and had signed a consent form. Some researchers found that repeated exposure to the same virtual environment with separation of less than seven days could significantly affect the levels of cybersickness which would induce participant's disorientation and nausea [13, 22]. Therefore, participants had not been exposed to the experimental VE in the previous 2 weeks.

2.2 Apparatus and the VE

The experimental environment was constructed by virtual developing software and presented on three types of displays: 3D monitor (with circular polarized glasses), HMD and 19" TFT-LCD display. The scene was designed as a retail store, which contained four subjective categories including stationeries, hand tools, cleaning articles and toiletries as shown in Figure 1. Stationeries and hand tools include

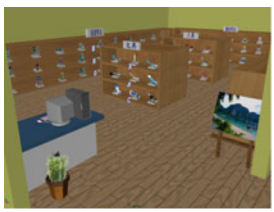
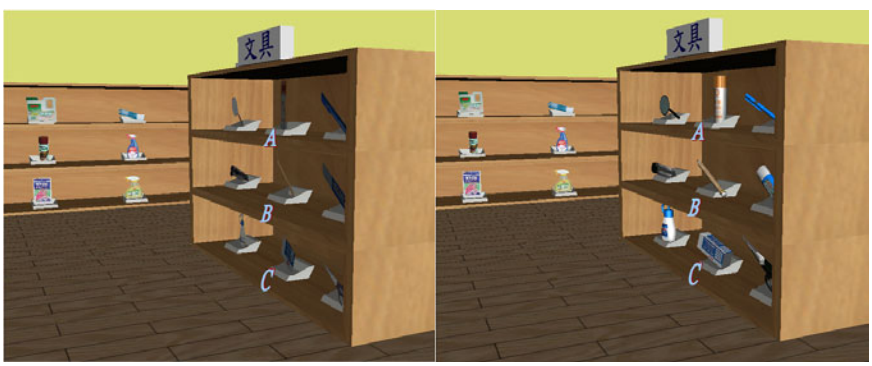


Fig. 1. A scene of the experimental 3D stationary store



(a) (b)

Fig. 2. A scene: (a) in low level depth cues with 2D images; (b) in high level depth cues with 3D stereopictures

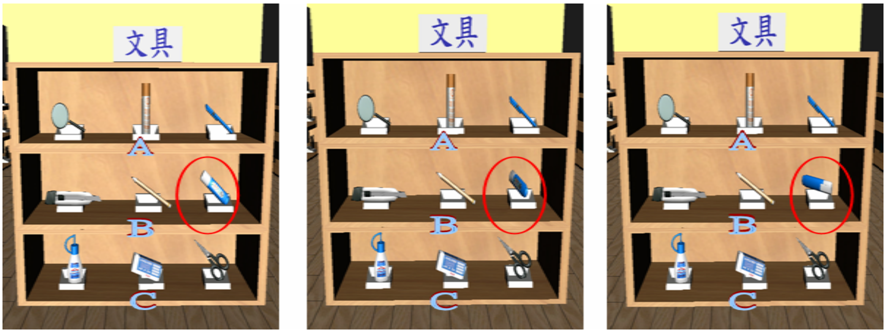


Fig. 3. The objects can be rotated along pitch and yaw axis in high level depth cues environment

eighteen goods exhibited respectively. Cleaning articles and toiletries include twenty seven goods exhibited respectively. Because this study was focused on the effect of presence in different depth perception cues, the scene was designed in two conditions: high and low level depth cues. Figure 2(a) shows a scene with low level depth cues, in which the goods are designed with 2D images. Figure 2(b) shows a scene with high level depth cues, in which the goods are designed with 3D stereopictures to appear to have good shape and depth. Additionally, the user can interact with objects by using the mouse. Figure 3 shows an object (a correction fluid) can be rotated along pitch and yaw axis for multiple viewpoints.

2.3 Experimental Design and Procedures

The study involved a 2 (level of depth cues: low and high) × 3 (mode of display: HMD, general display (22” TFT-LCD) and 3D monitor (Stereographic ready monitor with circular polarized glasses)), between-subjects experiment, resulting in a full-factorial design with six treatment conditions. Each participant was randomly assigned one of the six conditions to do the task of goods-finding. Therefore, there were ten participants randomly assigned to each one of the six conditions.

During the exposure period, participants were asked to search for and confirm some goods in the store. There were eight target goods need participants to search for, however, only six goods exhibited in the showroom, the others did not. When they found the target, s/he should write down the correct position on the check sheet (i.e. each showcase was numbered to let participant record number of position). If the participant found the target object did not exhibit in the showroom, s/he must remark “X” in the column of the object.

Before exposure, participants were asked to complete a Simulator Sickness Questionnaire (SSQ) documenting the severity levels of 16 sickness symptoms (Kennedy and Lane, 1993). SSQ is the most popular subjective measure for both simulator sickness and side effects experienced in VEs. If a participant reported any moderate symptom of discomfort or sickness in the pre-exposure SSQ, the participant was asked to rest for 10 min. prior to filling in a second pre-exposure SSQ. When all eight target goods had been found or confirmed not in the showroom, the experiment was completed. Finally, participants would be asked to complete Presence

Questionnaire (PQ) and SSQ. The PQ, devised to measure user presence within a virtual environment, consists of 32 questions regarding user interaction with 7-point scale rating (Lampton et al., 1994).

3 Results and Discussion

The initial statistical analysis revealed that the data set (presence scores and cybersickness scores with independent variables in different levels) was analyzed using the Shapiro-Wilk test to check the distribution of the sample. The results showed that the data set for each cell was normal distributed (P value > 0.05 in Shapiro-Wilk Test). Based on the results, an ANOVA analysis of presence scores and Cybersickness scores was performed. Based on the ANOVA results in Table 1and 2, it was apparent that the effects of depth perception cues and display modes are significant on presence and cybersickness symptoms in the elderly. More specifically, the following will be discussed.

3.1 Measures of Presence

Table 1 shows that depth perception cues in high level provided users sense of presence as having significantly higher than did in low level. It means that 3D stereopictures may have provided enough stereopsis and stereo acuity for users to closely check objects, examine objects from multiple viewpoints and well interact with the objects, thus users feel better presence than in 2D image of objects within the virtual store. In addition there is significant difference within display modes. Therefore, there is a need for further investigation into the different effects of display modes on presence. A Turkey's post-hoc test was used for pairwise comparison on display modes as shown in Table 3. The results show that the feeling of presence was the best on 3D monitor than others. It seems that when users browsed the virtual store

Table 1. ANOVA analysis of the effects of depth cues and display modes on presence scores

Sources		Means	SS	df	MS	F	P value
Depth cues	Low	76.8	14758.017	1	14758.017	72.244	.000*
	High	108.2					
Displays	TFT-LCD	79.2	11797.033	2	5898.517	28.875	.000*
	HMD	86.5					
	3D monitor	111.9					
Interaction			90.833	2	45.417	.222	.801
Error			11031.100	54	204.280		
Total			551237.000	60			

*p < 0.05 significance level

Table 2. ANOVA analysis of the effects of depth cues and displays on cybersickness scores

Sources		Means	SS	df	MS	F	P value
Depth cues	Low	13.5	784.238	1	784.238	26.419	.000*
	High	20.7					
Displays	TFT-LCD	13.3	534.793	2	267.396	9.008	.000*
	HMD	20.6					
	3D monitor	17.4					
Interaction			221.470	2	110.735	3.730	.030*
Error			1602.979	54	29.685		
Total			20645.698	60			

*p < 0.05 significance level

Table 3. Turkey's post-hoc tests for the effects of display modes on presence

(I) Display mode	(J) Display mode	Mean difference (I-J)	Std. Error	P value
HMD	TFT-LCD	7.250	4.520	.115
	3D monitor	-25.450*	4.520	.000
TFT-LCD	HMD	-7.250	4.520	.115
	3D monitor	-32.700*	4.520	.000
3D monitor	HMD	25.450*	4.520	.000
	TFT-LCD	32.700*	4.520	.000

*p < 0.05 significance level

with 3D monitor, the image will appear to have shape and depth and be, in a virtual sense, 3D. Then, the binocular disparity depth cues can be provided and benefit in making quick and accurate relative distance judgments may be proved. In advance, we found that through a combination of high level depth cues and 3D monitor as shown in Figure 4 (left side), phenomenon of stereopsis could be created easily, so the feeling of presence would be best for users. Although VEs were displayed via HMD that present separate images to create binocular disparity, the standard degree of binocular disparity was not sure to fit all users. So the average score of presence is fair. Additionally, browsing the virtual store in low level depth cues with TFT-LCD, the stereopsis could not created easily to limit depth perceiving, so the feel of presence would be poor.

3.2 Effects of Presence on Cybersickness

Participants' symptoms of cybersickness were evaluated by SSQ in the post-exposure. The results show that the effects of depth perception cues and display modes were significant. A Turkey's post-hoc test was used for pairwise comparison on display modes as shown in Table 4. It seems that no matter in low or high level depth cues, HMD would induce cybersickness easily. In addition the interaction effects between depth cues and display modes on cybersickness was significant, our concern becomes what mode of display with depth cues will result in significantly greater symptoms of cybersickness within a 3D virtual store. Figure 4 (right side) illustrates the interaction relationship between the effects of depth cues and display modes. It can be observed that the 3D monitor combined with low level depth perception cues would induce cybersickness slightly, but in high level depth perception cues participants suffered more severe cybersickness, even more serious than with HMD. It could be explained that the 3D virtual store with 3D monitor in high level depth perception cues would provide more in terms of interaction with objects, users might immerse in the VEs by

Table 4. Turkey's post-hoc tests for the effects of display modes on cybersickness

(I) Display mode	(J) Display mode	Mean difference (I-J)	Std. Error	P value
HMD	TFT-LCD	7.2930*	1.72293	.000
	3D monitor	3.1790	1.72293	.165
TFT-LCD	HMD	-7.2930*	1.72293	.000
	3D monitor	-4.1140	1.72293	.053
3D monitor	HMD	-3.1790	1.72293	.165
	TFT-LCD	4.1140	1.72293	.053

*p < 0.05 significance level

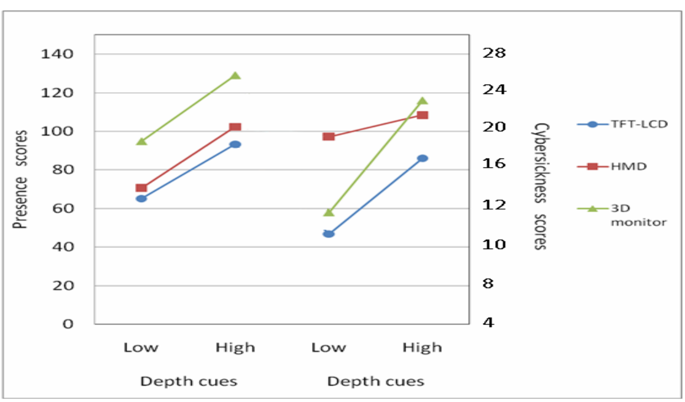


Fig. 4. Plots of mean scores of presence and cybersickness on depth cues for display modes (TFT-LCD, HMD and 3D monitor)

clicking and holding mouse button for moving, dragging or zooming in/out the goods and overall environment. Therefore, the symptoms of cybersickness are expected to occur for users due to speed and angle of the rotating scene and objects. Interestingly, in the same display instrument, there was a significant correlation between the experience of presence and incidence of sickness (r = 0.671, p = 0.000). It appears that the feeling presence adds to the incidence of cybersickness.

4 Conclusion

Virtual store with 3D images provides monocular depth cues to let older users experience good stereo acuity. The current study found that participants who have high level depth cues, and then do benefit from binocular disparity within a 3D monitor, were able to feel well presence. Although HMD could present separate images to create binocular disparity, the standard degree of binocular disparity was not sure to fit all users, so the feel of presence is fair. Additionally, there were significant differences found within the three display modes evaluated in this study in either reported sense of presence or cybersickness. Although the sense of presence in display with TFT-LCD was poor in either low or high level depth cues, the symptoms of cybersickness were the slightest. The older participants experienced presence so good in the high level depth perception cues of VE presented via a 3D monitor, but the symptoms of cybersickness was the most serious. In general, it is expected that exposure to 3D monitor will increase oculomotor disturbances due to the mismatched between the vection sensation and the vestibular cues under quick rotating speed, at a high angle of inclination for a prolonged period of exposure. In the future, it will be a step forward towards designing a warning system to detect improperly operating and prolonged exposure for combating cybersickness.

Acknowledgement. The authors would like to thank the National Science Council of the Republic of China for financially supporting this work under Contract No. 99-2221-E-238-014.

References

1 Cavoto, B.R., Cook, R.G.: The Contribution of Monocular Depth Cues to Scene Perception by Pigeons. Psychological Science 17, 628–634 (2006)
2 Cobb, S.V.G., Nichols, S., Ramsey, A., Wilson, J.R.: Virtual Reality-Induced Symptoms and Effects (VRISE). Presence: Teleoperators and Virtual Environments 8, 169–186 (1999)
3 Cutting, J.E.: How the eye measures reality and virtual reality. Behavior Research Method, Instruments, & Computers 29, 27–36 (1997)
4 Czaja, S., Lee, C.: Designing Computer Systems for Older Adults. The Human Computer Interaction Handbook 21, 413–427 (2003)
5 Freeman, J., Avons, S.E., Pearson, D.E., Ijsselsteijn, W.A.: Effects of sensory information and prior experience on direct subjective ratings of presence. Presence: Teleoperators and Virtual Environments 8, 1–13 (1999)

6 Hale, K.S., Stanney, K.S.: Effects of low stereo acuity on performance, presence and sickness within a virtual environment. Applied Ergonomics 37, 329–339 (2006)
7 Howarth, P.A.: Empirical studies of accommodation, convergence, and HMD use. In: Proceedings of the HOSO-Bunka Foundation Symposium, Tokyo (1996)
8 Ijsselsteijn, W., Ridder, H., Hamberg, R., Bouwhius, D., Freeman, J.: Perceived depth and the feeling of presence in 3DTV. Displays 18, 208–214 (1998)
9 Johnson, R., Kent, S.: In: Designing universal access: web-applications for the elderly and disabled, vol. 9, pp. 209–218 (2007)
10 Jones, S.: Generations online in 2009. Pew Internet & American Life Project Report (2009)
11 Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. International Journal Aviation Psychology 3, 203–220 (1993)
12 Lampton, D.R., Knerr, B.W., Goldberg, S.L., Bliss, J.P., Moshell, L.M., Blau, B.S.: The virtual environment performance assessment battery (VEPAB): development and evaluation. Army Research Institute, Orlando (1994)
13 Lathan, R.: Tutorial: a Brief Introduction to Simulation Sickness and Motion Programming. Real Time Graphics 9, 3–5 (2001)
14 LaViola, J.J.: A discussion of cybersickness in virtual environment. SIGCHI Bulletin 32, 47–56 (2000)
15 Lombard, M., Ditton, T.: At the heart of it all: the concept of presence. Journal of Computer Mediated Communication 3 (1997), http://jcmc.indiana.edu/vol3/issue2/lombard.html
16 Mikropoulos, T.A., Strouboulis, V.: Factors that influence presence in educational virtual environments. Cyberpsychol. Behav. 7, 582–591 (2004)
17 Nichols, S., Haldane, C., Wilson, J.R.: Measurement of presence and its consequences in virtual environments. International Journal of Human-Computer Studies 52, 471–491 (2000)
18 Sadowski, W., Stanney, K.M.: Presence in virtual environments. In: Stanney, K.M. (ed.) Handbook of Virtual Environments: Design, Implementation, and Applications, pp. 791–806. Lawrence Erlbaum Associates, Mahwah (2003)
19 Sheridan, T.B.: Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments 1, 120–125 (1992)
20 Spek, V.D.: The effect of cybersickness on the affective appraisal of virtual environments. Master's thesis, University of Utrecht (2007)
21 Stanney, K.M., Mourant, R.R., Kennedy, R.S.: Human factors issues in virtual environment: A review of the literature. Presence 7, 327–351 (1998)
22 Stanney, K.M.: Handbook of Virtual Environments. Earlbaum, New York (2002)
23 Steuer, J.: Defining virtual reality: dimensions determining telepresence. Journal of Communications 42, 73–93 (1992)
24 Sylaiou, S., Karoulis, A., Stavropoulos, Y., Patias, P.: Presence-centered assessment of virtual museums' technologies. Journal of Library and Information Technology 28, 55–62 (2008)
25 Wickens, C.D., Todd, S., Seidler, K.: Three-Dimensional Displays: Perception, Implementation, Applications, Crew System Ergonomics Information Analysis Center (1989)
26 Wilson, J.R.: Autonomy, interaction and presence. Presence: Teleoperators and Virtual Environments 1, 127–132 (1997)
27 Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence 7, 225–240 (1998)

A Development of Web-Based Player for Instructions Recorded with the Electronic Blackboard System IMPRESSION

Yuichi Ohkawa and Takashi Mitsuishi

Graduate School of Educational Informatics, Tohoku University
27-1 Kawauchi, Aoba-ku, Sendai, Miyagi 9808576, Japan
{kuri,takashi}@ei.tohoku.ac.jp

Abstract. In these years, we have been developing the interactive and sharable electronic blackboard system; IMPRESSION. In this paper, we described the development of a web-based player for instructions, which are performed with IMPRESSION to conduct a lesson. By using this system, a log, which shows performed instructions in a lesson, can be used as an on-demand e-learning material. In addition, this system provides various operational functions for playback instructions and also can record learners' actions of studying with these instructions. As a result, it makes possible to a teacher to grasp learners' understanding.

Keywords: E-Learning, Electronic Blackboard, RIA, User Interface.

1 Introduction

In order to conduct a flexible and effective lesson, many teachers try to modify their instructional plan based on formative evaluation in lesson. Moreover, they review their lesson as reflection after the lesson. In order to help such activities, we have been developing the interactive sharable electronic blackboard system IMPRESSION [1] these years [3,4,5]. However, this system has developed as an application on Windows OS. Therefore, users of this system require administrator privilege of client PCs or technical supports by administrators. It brings obstruction of easy use for user of this system.

In addition, because this system is aimed to teach by modifying instructional plan instantly based on formative evaluation in a lesson, it is not considering to use instructional records, which are log data of teaching with IMPRESSION, as on-demand e-learning materials.

On the other hand, various rich internet applications (RIA) using the technology such as Adobe's Flash and Microsoft's Silverlight have been developed and utilized in recent years. A RIA has greater interactivity compared with traditional web applications by using JavaScript, and can be executed in various OS. Therefore, we think that the learning material, which is developed by using such RIA, can achieve the learning environment with high potential of expression and interactivity accompanied with the learner's action. If learners could take various actions (e.g.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 500–509, 2011.

taking notes, bookmarking and seeking of a learning material) easily according to their own understanding, a high learning effectiveness could be expected even if teacher did not grasp learners' understanding. Moreover, a learning environment using RIA can record learners' action attending class in detail. Then, if the recoded actions were analyzed and fed back to learners with teacher's knowledge, an effective learning can be expected.

In this paper, we develop such a player with Flash technology, which can play back instructions recorded from IMPRESSION and also records learner's actions for playback. By using Flash, it does not require complex operations to install a program and is able to execute it on the most of Web browsers.

2 Electronic Blackboard System IMPRESSION

2.1 Outline of IMPRESSION

The IMPRESSION [1] (Interactive Multimedia PREsentation System for Shared Instructional Objects on the Networks) is an electronic blackboard system targeting both synchronized education such as distance education relayed over network and face-to-face education in classroom, based on the “double loop” instructional design process model [2].

A double loop model is aimed at teaching interactively. As shown in Fig. 1, this model has an external loop that means a flow of a plan, application and evaluation of instructions, and also have an internal loop which provides a flexible modification of instructional plan based on a formative evaluation, means a flow of an implementation, check, modification in a lesson.

The original IMPRESSION developed in 2006 by Higuch *et al.* has functions for the internal loop of double loop model. For example, if a learner indicated a question directly or over network when he/she feel uncertain about something in a lesson, teacher can explain repeatedly with this system using a new slide [1].

Based on the original system, Suzuki *et al.* [3] extended this system in 2008, and a teacher is able to plan an instruction plan in a tree structure and conduct a lesson with a new IMPRESSION (Fig. 2). In 2008, Kanno *et al.* also extended some functions to record performed instructional activities in a lesson and play back them after the lesson [4] in order to facilitate teachers for their reflection (Fig. 3).

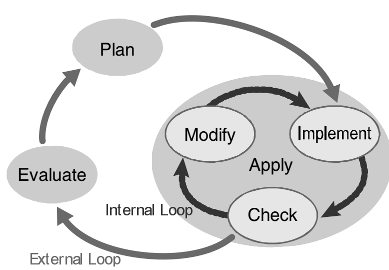


Fig. 1. Double loop model

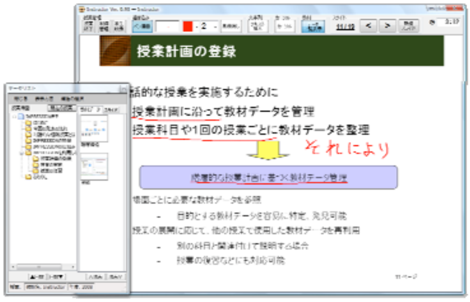


Fig. 2. An example of execution of the IMPRESSION

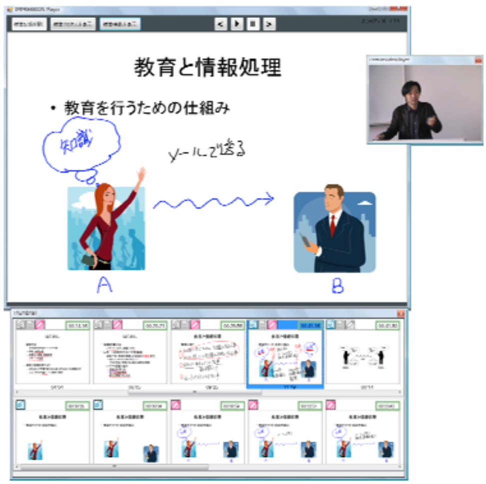


Fig. 3. The IMPRESSION for teacher reflection

2.2 Additional Requirements for the Current IMPRESSION

Requirements for On-demand e-learning Materials. Generally, when we provide lectures for e-learning, we should conduct lessons in real time as synchronous distant lectures or provide stored lectures on a server by on-demand. But, because the IMPRESSION, we have developed, is intended to conduct a lesson synchronously, asynchronous feature of the IMPRESSION was only playback function for teacher reflection. And, when we want use the playback function, it requires extra application program.

However, asynchronous e-learning has also advantages that learners can decide when to learn by themselves. Therefore, recorded instructions of synchronous lessons with the IMPRESSION have great potential to be used as asynchronous e-learning materials.

On the other hand, asynchronous e-learning have limitation that teacher cannot manage progress of learners' learning, and learners should manage his own learning. However, existing asynchronous e-learning systems are restricted so much for learners to control their learning progress. For example, when they provide learning materials consisting of slides and video, they have only limited functions like temporal pause while playing, zapping slides and so on. And, when learning materials are consisting of only video, they don't provide more than a pause and seek. There are few systems that can find a scene where learner would like to play back easily. As a result, because of the difficulty of operation of the system, learners leave questions lay although they have some questions about learning materials. It might cause that learners cannot understand enough.

For teacher reflection, the IMPRESSION has functions to find a scene from some slides and events like teacher's handwriting on blackboard. However, a teacher cannot rewind and check a preceding moment of video in order to know a reason why teacher wrote the explanation on blackboard. So, some teachers who used the IMPRESSION are requesting improvement for such functions.

We estimate so many learners will feel unmotivated to learn with e-learning materials because of such problem. In order to solve such problems, we believe a more flexible learning environment is required.

Requirements for Implementation Environments. As described in section 0, in order to execute the current IMPRESSION, the system must be installed on Windows OS as a native application. As a result, following problems are identified; (1) Many people are difficult to use this system because of trouble to install the application. (2) Other platforms such as Mac OS and Linux are not supported. (3) Frequent updates and flexible improvement with change of networking protocol are difficult due to the required re-installation.

That is, in order to conduct a distance lesson with the IMPRESSION, it requires install application into not only a teacher's computer but also learners' computers. When learners were not familiar with PC, it is difficult for a teacher to support individual learners, and it might disturb a distance lesson.

When learners are using Mac OS, Linux, or some other OS, it also might cause same problem.

In addition, it is difficult for a teacher to manage revision of installed software on learners' PCs, and it might prevent improvement of the IMPRESSION (e.g. change of networking protocol.) In such cases, learners might be not able to watch learning materials at worst.

In order to solve these problems, we decide to develop the web-based learning system with Flash technology, which can play back instructions of a conducted lesson by using the IMPRESSION.

3 Design of Asynchronous Learning System with Flash

3.1 Required Control Functions for Learning by Learners

As described above, a teacher cannot manage learning progress of learners according to their understanding when they are learning with asynchronous e-learning materials. Only learners can manage their learning progress by themselves. Thus, if learners could operate learning system more easily according to their own understanding, their learning must become more effective.

So, we propose to provide functions such as a fast forward, fast rewind and so on by the web-based player for IMPRESSION as well as standard playback functions like play and pause (Table 1). In this table (a) and (b) are functions to play and to pause a learning material, which are available in standard video player. The function (c) aims at checking a learning material in a short time by playing at high speed (2 times or 4 times as fast as the normal). Both (d) and (e) are functions for playing backward, which can be used to check a learning material while rewinding. And, the function of (e) can be used for high-speed playback as same as (c) except backward direction. The function of (f) can be used to find a scene where a learner would like to play back by using a slider. The function of (g) can be used to move at fixed time forward or backward. Finally, the (h) can be used to move to the scene while making the thumbnail of slides a clue.

These functions aim at learners' easy control of his own learning. For example, learners are able to seek important scene and watch it repeatedly, or skip unimportant scene.

Table 1. List of available functions in the proposed system

(a)	Play	(e)	Fast rewind (x2, x4)
(b)	Pause	(f)	Seek by using slider
(c)	Fast Forward (x2, x4)	(g)	Skip before or after 30sec
(d)	Reverse play	(h)	Jump to each slide with thumbnail

3.2 Potential of Rich Internet Application Using Flash Technology

This time, in order to implement the above functions on various environment, we developed our system as a rich internet application (RIA) by using the Flash technology. The Flash plug-ins have been installed into most of web browsers usually as standard, and they enables applications to perfume interactively as well as native applications.

In addition, by providing the system as a web-based application it is not necessary for learners to install and update the application and they can use it, even if the system revision has been updated.

3.3 A Function for Logging of Learners' Behavior

In this study we also intend to develop functions to record learners' behavior. If detailed actions of learners, who are using various functions to play back recorded instructions as described above, are acquired, a teacher might be able to grasp condition of learners. On the other hand, if behavior of learners was analyzed and fed back to himself/herself, it might be help for him/her to understand his/her own condition and to learn effectively.

Fig. 4 shows an example of the executed the prototype system we have implemented. This prototype system consists of 4 units and 2 external servers; a blackboard unit in which displays pictures and strokes annotations, a control unit to specify playback direction and speed, a video unit in which video of lesson is displayed, and a behavior recording unit by which behavior of lectures are recorded. Fig. 5 shows the system architecture of the implemented system.

Fig. 4. A prototype of proposed system

4.1 Blackboard Unit

This unit has functions to interpret the instructional data generated by the IMPRESSION and to play back pictures, drawing handwriting annotations, and so on.

An instructional data generated by the IMPRESSION forms in XML, and every instructional actions are recorded with their time code as shown in Fig. 6. Therefore, the proposing system parses the data and displays instructions according to time codes.

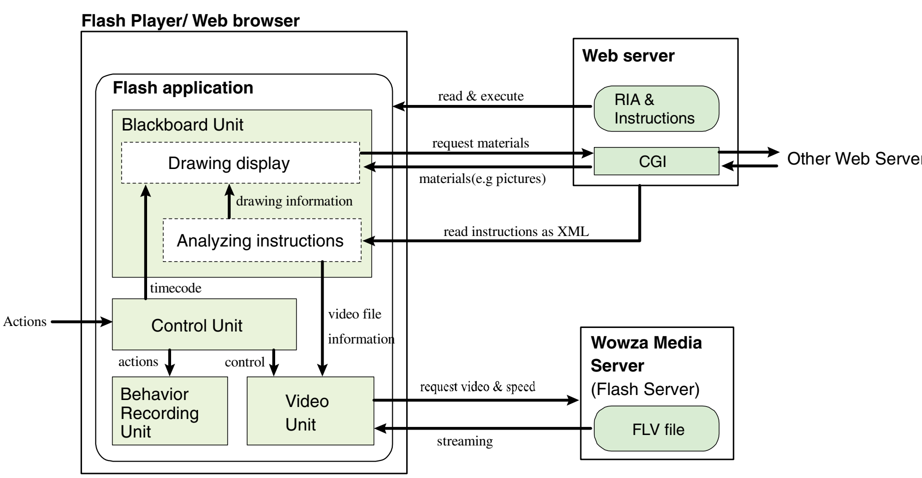


Fig. 5. A system architecture of proposed system

The previous IMPRESSION had a problem that it represented handwriting strokes for annotation as bitmap picture. Because the previous IMPRESSION was developed to use in real time during a lesson, and only playback of forward direction was assumed. So, when an annotated bitmap on a slide, which was already used once, need to be shown again, we should only store last condition of annotated bitmap and display the stored bitmap with the slide. On the other hand, the system that we are implementing has a function of playback in backward direction. Annotated bitmaps that are overwrote or erased before must be restored in backward direction. Therefore, in proposed system all operations for annotation such as drawing line are expressed as independent drawing objects including operation of erasing line. And, each drawing objects are displayed and hidden according to the time code.

```xml
<?xml version="1.0" encoding="euc-jp" ?>
<lecture>
  <capture name="080415.wmv" offset="+12"/>
  <material-list>
    <define name="Image1" url="http://www.ei.tohoku.ac.jp/lecture/01/communication-system-model.png" />
    <define name="Image2" url="http://www.ei.tohoku.ac.jp/lecture/01/computer-network-system-model.png" />
    <define name="Slide1" url="http://www.ei.tohoku.ac.jp/lecture/01/slides/slide-01.png" />
    <define name="Slide2" url="http://www.ei.tohoku.ac.jp/lecture/01/slides/slide-02.png" />
  </material-list>
  <start date="2008/04/15/13:05:35" />
  <slide time="504" operator="INSTRUCTOR(002)-takashi" >
    <add />
  </slide>
  <operate time="560" operator="INSTRUCTOR(002)-takashi" >
    <present name="Image3-1" x="-112" y="-140" />
  </operate>
  <operate time="581" operator="INSTRUCTOR(002)-takashi" >
    <change name="Image3-1">
      <form x="70" y="45" width="989" height="617" />
    </change>
  </operate>
  <draw time="594" operator="INSTRUCTOR(002)-takashi" >
    <line name="blackboard" color="#FFFF0000" width="2">
      <point x="278" y="231" />
      <point x="277" y="224" />
      <point x="277" y="225" />
      <point x="278" y="226" />
      <point x="279" y="228" />
      <point x="280" y="230" />
      ....
```

Fig. 6. An example of recorded instructions in XML

4.2 Control Unit

The control unit consists of 3 subunits; a main controller ((1) in Fig. 4), a thumbnail panel (2), and status display (3). The main controller has 7 buttons and 1 seek bar to control playback. The thumbnail panel has thumbnail of slides. And, the status display shows time code and status of playback.

User interfaces for the functions (a) to (g) shows in section 3.1 are implemented as shown in Fig. 7. The functions fast forward and backward with double speed or quad speed are toggled by clicking fast forward/backward button. When the playback speed is changed, triangle marks will be shown in the status display according to the direction and speed.

Fig. 7. The interface of main controller

The function (h) in section 3.1 is implemented by displaying thumbnails on the left side of screen as shown in Fig. 4. If one of these thumbnails was clicked, playback jumps to the slide. Here, each thumbnail are showing final state of explanation with the slide not beginning of that, because beginning state of slide in the IMPRESSION is almost blank and a teacher annotates something over the slide.

In order to avoid unusefulness of narrow window area of web browser, the user interfaces of the main controller and the status display can be moved to any place in window. And, these user interfaces can be made transparent by using slider upper right each user interfaces as shown in Fig. 8. In addition, the slider and the frame of these interfaces will be hidden automatically in a few seconds after focus of mouse on these interfaces comes off as shown in Fig. 9. The interface of the thumbnail panel is also visible only while getting a mouse focus.

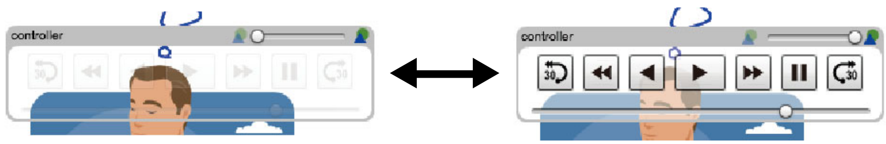


Fig. 8. A transparency control of interfaces

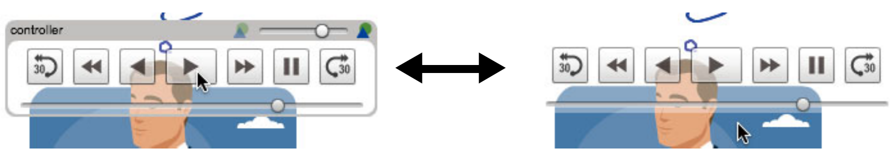


Fig. 9. An example of automatic hidden control according to focus of mouse

4.3 Video Unit

The extension by Kanno *et al.* [4,5], the existing IMPRESSION came to be able to record a video of classes with the instructions such as an annotation. In this study, we implemented to play a video recorded by the IMPRESSION according to time code of the instructions. However, because the IMPRESSION records a video as a Windows Media format, it cannot be utilizable in Flash applications. Therefore, we trans-coded a video into FLV format.

In this study, we used the Wowza Media Server that is one of Flash servers to deliver the function to change the speed of playback. However, we found the problem that the audio of video cannot be played when the speed of playback was changed.

In addition, the transparency of interface of the video unit also can be controlled, and a part of interfaces is hidden automatically according to focus of mouse as well as controller unit.

4.4 Behavior Recording Unit

With the implemented system, it is possible to record learner's behavior when and which operation such as fast forward does each learner do with the user interfaces, and recorded behavior can be confirmed from context menus as shown in Fig. 10.

At this moment, we have not implemented the functions to collect and analyze these records. We think it is necessary to discuss implement such function in the future.

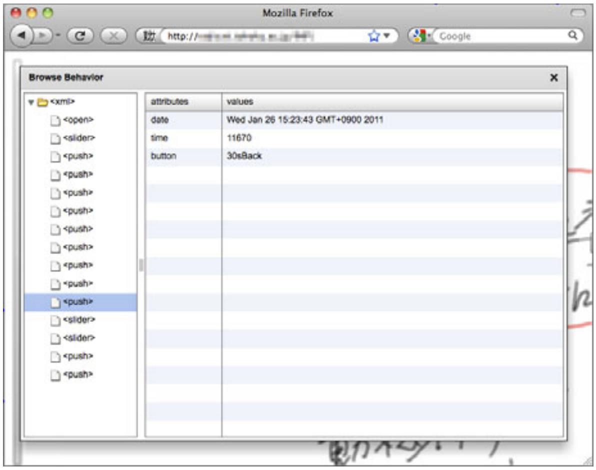


Fig. 10. An example of checking learner's behavior

4.5 Countermeasure for Security Policy of Flash

The existing IMPRESSION allows to use various multimedia data provided by public web servers on the Internet as educational material, and each of terminals of the IMPRESSION downloads these data directly from such public web servers in order to reduce communication load between teacher's terminal and learners' terminals.

However, the security policy for current version of the Flash Player prohibits downloading data directly from the other web servers. Therefore, we prepared a proxy server for the implemented system to download such data through the proxy server as shown in Fig. 5.

5 Conclusion

In this paper, we proposed a web-based IMPRESSION player which can play back the instructions as a result of the conducted lesson by using the IMPRESSION. By using the implemented system, we are able to reuse synchronous lessons conducted with the IMPRESSION as asynchronous educational material for e-learning, and described design and implementation of it.

However, we have not investigated the effectiveness of our system. We think it is necessary to perform some practical experiments in near future. And, we also need to discuss the functions to analyze recorded logs of learners' behavior with our system.

Acknowledgement. This work was supported by Grants-in-Aid for Young Scientists (21700798 and 20700631).

References

1. Higuchi, Y., Mitsuishi, T., Go, K.: An Interactive Multimedia Instruction System: IMPRESSION for Double Loop Instructional Design Process Model. IEICE Transaction of Information & Systems E89-D(6), 1877–1884 (2006)
2. Higuchi, Y., Konno, F., Mitsuishi, T., Go, K.: A Double Loop Instructional Design Process Model for Interactive Instructions. Japan Journal of Educational Technology 31(4), 457–468 (2008)
3. Suzuki, T., Konno, F., Ohkawa, Y., Mitsuishi, T.: Design and Implementation of a Lesson Plan Based Data Management Mechanism for Educational Materials in IMPRESSION. In: Proceedings of 33rd Conference of Japan Society for Information and Systems in Education, pp. 322–323 (2008) (in Japanese)
4. Kanno, Y., Konno, F., Ohkawa, Y., Hashimoto, K., Mitsuishi, T.: A Proposal of Teacher Reflection Supporting System Based on Observation of Reviewing Activities with Handouts and Video Records. Educational Informatics Research, pp.1–8 (2008) (in Japanese)
5. Konno, F., Kanno, Y., Ohkawa, Y., Mitsuishi, T., Hashimoto, K.: Design and Implementation of a Teacher Reflection Support System using IMPRESSION and MidField. In: Proceedings of ED-MEDIA 2009, pp. 2421–2430. AACE (2009)

Categorize Web Sites Based on Design Issues

Amin Rasooli[1,2], Fattaneh Taghiyareh[1], and Peter Forbrig[2]

[1] School of Electrical and Computer engineering, University of Tehran, Tehran, Iran
[2] Department of Computer Science, University of Rostock, 18051 Rostock, Germany
amin.rasooli@uni-rostock.de, ftaghiyar@ut.ac.ir,
peter.forbrig@uni-rostock.de

Abstract. Interface design is one of the most important topics during web development process. The final design is a tradeoff between the owner's personal idea and the web developer's perception of what he wants. In this paper, we have proposed a new model called WLDM (Web layout design model) to cover the important components of interface design. There are three components in the WLDM, including structure, content and visual. We have selected three features for structure, two for content, and three for visual component. Thereafter, we have made a dataset using 1088 most visited web sites. Finally, applying K-means algorithm, we have clustered this dataset. According to our result, six clusters were identified. Considering WLDM, web layout designer have a blueprint to cover areas of research related to this issue. The result of this clustering can be used for recommender systems to map owner groups, which have different attitude.

Keywords: WLDM, Web design, Adaptation, Personalization, Layout, User Interface, Human-Computer Interaction, Clustering, Data mining.

1 Introduction

Web design is the task of making presentation of content, which are delivered to end users through the Internet network by using a Web browser or other Web based software. "About 40% to 70% of a web site's code is its user interface".[1] The goal of web design is to make a collection of some documents and elements that put on a web server or more, and illustrate content and user interfaces to the visitors in form of some hypertexts. Some items like text, images, and forms can be put on the page using HTML tags. Displaying more complex media like graphics graphs, animations, videos, and sounds needs some plug-ins such as Adobe Flash, Quick Time and so on. [2], [3]

Despite the varieties and volumes of web site templates [4], the owners are looking for a more principled and attractive layout for their web site rather than other web sites. In this situation, each part of webpage design can have an essential role in the success of a web site, but this process is not always efficient and adjustable for maximize usability of the web site; it may be inefficient and will be totally different to his taste.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 510–519, 2011.

Among the reasons for aforesaid condition, we can mention the incompatibility between the web site owner’s needs and the final implemented web site. In other words, the problem is lack of the web site manufacturer's understanding of owner's needs and his visitors’ needs the conditions described by owner. This phenomenon leads to making a product, which do not satisfy anyone. This kind of problem is usually happens to the owner who does not have enough information about the web design, or sometimes he cannot present what he exactly wants (actually it results from the lack of the owner's knowledge about web site or the lack of designer's familiarity with situation). Therefore, if an owner knows exactly what he wants, then designing companies also understand his expectation, then they can help him to have a suitable web site. Otherwise, they will encounter many problems either in the initial design or in delivery of the final product. [5]

We started by introduction of our aims then present model of web interface features in Section 2. In Section 3, we did clustering with K-mean algorithm on 1088 web sites. Finally, we will summarize contributions with suggestions for future research in Section 4.

2 WLDM Model

Our proposed model is for supporting and adapting design personalization that we called it as WLDM (Web layout design model). In the following, we can see the usage of this model in order to improve design of the web site’s layout. [6]

Different components of the web site are mentioned, as you can see in Fig. 1. It can be divided into three different categories: structure, content, and visual. We will discuss in details about these three components and their branches, in follows. [7][8]

2.1 Structure

Frame and formation of a web site is related to structure components. The structure of web site made up from different parts, we consider these items: types of web interface, links and cognitive load.[9]

Types of web interface. Web sites can be categorized in different types. Each one is specialized in a particular service or usage. Usability, visualization, and functionality are three distinct issues, meanwhile they related on type of web site. Six major types according to some researches which, are Graphical, Menu driven, Search engine type, Discussion board type, Content managed web site and Flash web site.[10]

Links. We analyses links from two standpoints, on one hand place and appearance of links, And on the other hand navigation and target issues.

Replacing links in a prominent place for example on the main body of web site and putting similar items close to each other. Only have unique navigation system, not different way for specific path. Using icons is an option for representing links in web site but it is better to considering relation between links and the images. [11]

Distinction of links in a web site is the idea that based on some links lead to general information and some of them lead to detailed information. It should be considered that if it is better to put all the content (for example for news) or only a link for head line. [12]

Finding the intended item for the left presentation position was appreciably faster than that of the bottom and top positions. In addition, for the right position found to be notably faster than the bottom position. Menu items should be placed vertically in order to decrease speed of finding a word. [13] And horizontal menus are searched more quickly than vertical ones. [14][15][16]

Cognitive load. A common problem for computer users is Learning and remembering the information which is presented on a computer screen. Based on cognitive theories, one of the reasons for lacking the retention is user's inability to create a mental schema, picture, or plan about the information is illustrated on a computer screen. To create a schema or plan, users should find out, which an obtained knowledge is related on a big picture.

If you are lost on a web site several times, there must be a problem and confusing informing a schema so it should considered seriously for better retention by users. We can conclude that more information cannot cause more cognition, the ways of providing the information can make a schema, shape, or plan for better understanding better. [17]

2.1 Content

In general, "content" contains everything in your web. This may involves documents, data, applications, service. Moreover, if we had determined that a web site had not completed by the qualified stuff, we should logically progress to evaluating the properness of web sites. [9] Four issues of content investigation is as follows. [18]

Link Suggestion. "Reduce the number of links to ensure that genuine and necessary links are clearly identified". [12]

Cognitive load and impairment of learning can be the result of higher number of links for hypertexts per page. Several links as a navigation support techniques are recommended for leading learners and reducing cognitive overload practically. It is recommended to use hypertext design for accessing the information in an easy way and can help to improve user understanding on cognitive issues and so on can cause increasing in learning.[19]

Hypertext design is directed for enabling information access in a simple way and leading to an improvement of readers' understanding of cognitive stuff and increasing in learning.

Content faults. When we redesign or change the web site, it can cause some faults which should be considered and eliminated. Finding faults and fixing them can be classified according to type of service, failure identifiers, fault analyzers, analysis and repair tools. For more information we can discuss about the some of them as following;

- Failure identifiers (some failures could be discovered by user actions simulation, for example filling a form).
- Fault analyzers (it found out the failures and highlighted them).
- Analysis and repair tools (they help fixing the faults). [20][21]

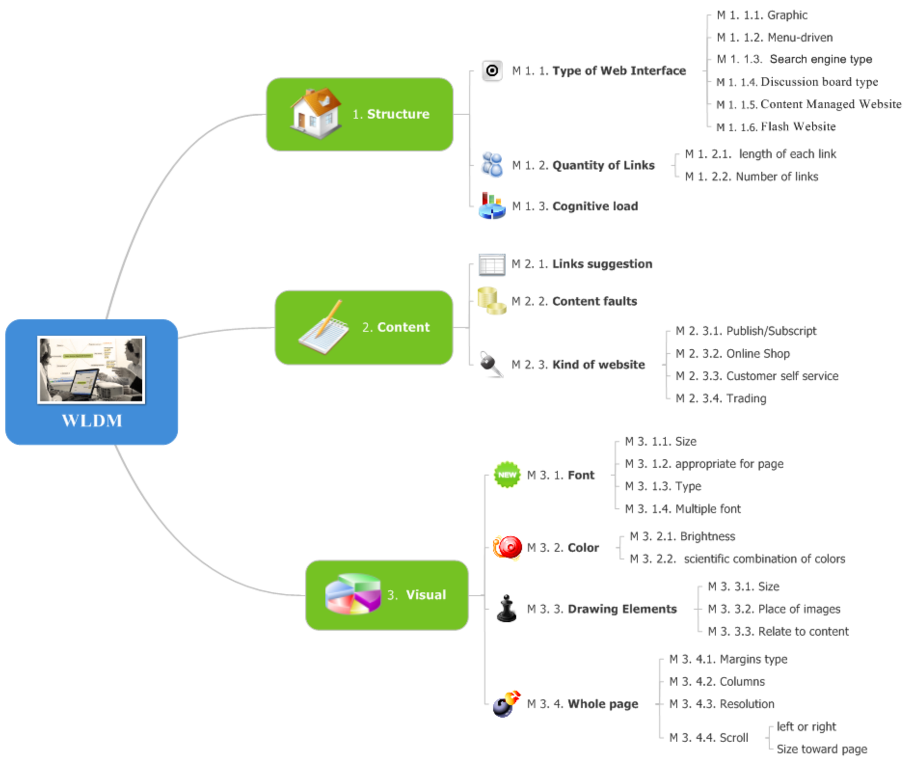


Fig. 1. Personalization Model of important components in a web site

Kinds of web site. In our study, we used the web site's type, which adopted the compact IBM classification of Web sites according to volume of traffic. Based on criteria such as: pages retrieved, number of transactions, their complexity, type, and number of searches, information stability, and security concerns, this classification proposed five types of high-volume Web sites: publish/subscribe, online shopping, customer self-service, trading, and B2B. [1]

2.3 Visual

Visual items aim to suggest a basic concept and prototype. If users want to experience personalized page design, they should express their preference about page design elements such as font, color, drawing elements and the whole page. Then, the personalization system simply assembles design elements selected by the user into a webpage and displays the result. We consider the element in detail.

Font. A significant element in designing webpage is the font and its attributes like size, type and so on. These attributes are very important because of their affect on the speed of reading and could attractive users. Some significant differences in reading efficiency were detected between the font types.

In general, Times and Arial type are read faster than Courier, Schoolbook, and Georgia. Fonts at the 12-point size read faster than fonts at smaller or bigger size. For font attractiveness, Georgia is recognized as being more appealing than Arial, Courier, and Comic, while Times is recognized as more appealing than Courier. [22]

Generally, older adults are very sensitive to font size on a page. The senior adults were changed with performing tasks such as finding special information on a page. [23]

Color. Color is one of the most worthy items in psychological aspect in this research area. There is a close relationship between colors of web sites and emotions. Web site colors can effect on Users' behaviors and cognitive processes especially when they are associated with a specific physical layout. A few recommendations are useful for appropriate color selection for designing user interfaces. It could be comprises of some aspects of interfaces like text background. In visual aspect, harmony is related on a suitable arrangement of different parts and some while it is enjoying by the eyes. In a web site, people should feel a sense of equilibrium and pleasure.

Color has different meaning in different culture and different nation have different expectation, and belief based of the meaning of each colors. So each color has its own psychological influences in any web site.

Drawing elements. Drawing elements can affect the web site in different aspects. Using pictures with high resolution increase the size of the web sites. Then for each picture, appropriate format and size is needed. Pictures and icons should be in harmony with the web site's type, background and other details of the web site.

Meanwhile picture's brightness, place of pictures and being related to the content are the important criteria should be considered.

Whole page. Total view, web site's margin, diagrams, and resolution are some of the parameters in this area. Diagrams and columns should be symmetrical otherwise, it is important to follow specific rules. The web site's resolution depends on the user's browsers and should get regulated with them.

Another point that should be taken into the consideration is scroll. Moreover, it is required not to use vertical scroll in the web site. It is often cited as evidence that increasing line spacing decreases reading time. This finding is attributed to the fact that close spacing of text requires more "eye fixations per line" and, as a result, fewer words are read during each eye fixation. [24]

3 Data Gathering from Web Sites and Clustering

One of the aims of this article is to cluster the web sites based on the mentioned features, which is selected by considering the eight important issues from all of them. These features are:

- Cognitive load: It would measure by sum of the lines and words in the homepage.
- Type of web interface: It consists of six types but we consider three of them such as normal html, graphical (or flash) type, and the other types. These types investigate by detecting words in introduction of the web site.

- Complexity: It related to the time of decoding and compiling the web site in some parts by our software. "a" consider as the low complexity, "b" as middle, and "c" as the high complexity.
- The quantity of the Links: Number of the links in the homepage.
- The quantity of the Images: Number of the images in the homepage.
- Web site's goal: It has got four options: Publish/Subscript, Online Shopping, Customer self-service, and Trading. They have been recognized with analogy issues, selecting, and mapping significant words in the web site.
- Font variety: font changing quantity.
- Color variety: color changing quantity especially in text.

In Table. 1, We can see the parts of the input of the program.

Table 1. Parts of Automatic Production of Web sites' Attributes

Brand	Cognitive load	Comp lexity	Type of WI	links	images	Web site's goal	Font	Color
google.com	741	a	HTML	17	6	Customer	11	47
live.com	3054	b	HTML	51	16	Publish	3	7
yahoo.com	7319	b	etc.	53	23	Online	14	162
microsoft.com	3596	b	HTML	71	15	Customer	2	0
facebook.com	1498	a	HTML	14	5	Customer	0	0
youtube.com	1201	b	Graphical	39	75	Publish	0	2
wikipedia.com	2768	a	HTML	33	51	Publish	10	3
aol.com	3463	c	etc.	49	45	Online	3	7
ebay.com	1757	b	Graphical	89	91	Online	-1	5

The software, which is used for extracting these data, has been written in PHP [25]. We investigated more than thousand of the popular web sites [26][27]. Next step was cleaning and preprocessing the data, and after preparing data, we used a popular data mining software, which is called Weka 3.6.3. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality [28]. We used the K-means method. Consequently, web sites are divided into 6 clusters [29].

The result of Web site clustering is exhibited in Table .2. These web sites were clustered according to the above-mentioned features. We had considered K between 2 till 20 as the quantity of the clusters by k-means algorithm. For four of the clusters we got optimum result.

The result of the clustering is illustrated in Table 2. Missing values were replaced with mean/mode. Within the clusters, the sum of squared errors is 367.83504. In addition, Fig 2 illustrate the vector of sum of squared errors by different amounts for k. Regarding to this graph (Fig. 1) we can see the largest drop in the sum of the within sum of squares in k=4 as the best quantity number of our clusters.

Table 2. Clusters centroids

Attribute	Full Data (1088)	0 30%(352)	1 40%(432	2 20%(220)	3 10% (108)
Cognitive load	63423.5441	110945.378	41849.9167	36468.5273	60300.8519
Type of WI	html	graphical	Html	graphical	html
Complexity	simple	complicated	simple	Complicated	complicated
Quantity of links	28.5625	56.54756	19.94352	6.74	22.5
Quantity of images	C	c	a	b	a
Web site goal	Trading	Online	Online	Online	Customer
Font variety	1/4	1/4	1/4	1/4	7/10
Color variety	21.4853	35.1585	14.5278	9.5818	32.037

This part of this study can be used as follows; first of all the owner will be classified, then each group will be mapped to our clusters as a class column, which is mentioned in Table 2 [30]. Finally, a new model is created. When new owner wants a layout, we can map him to one of the owner's group. The methodology and results help identify new layout by using collaborative filtering in recommender systems for the new owner [31][32].

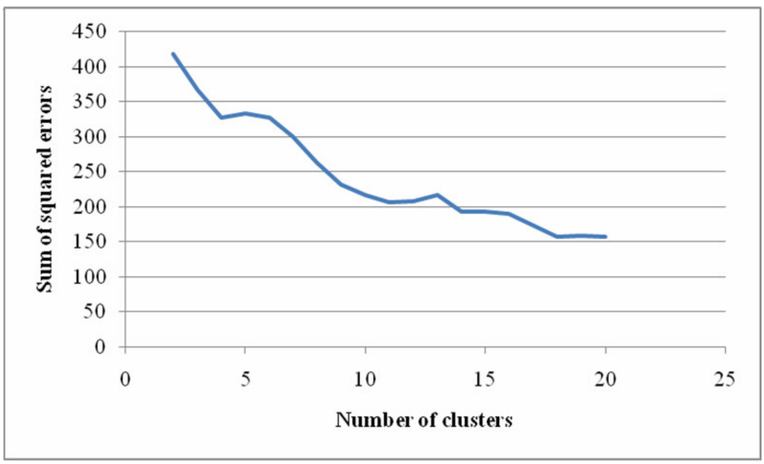


Fig. 2. Sum of squared errors by different amounts for k (between 2 until 20)

4 Future Works and Conclusion

This study comprises of two stages, the first stage was collecting the information about the important issues in web designing, and the result has been presented as a model named WLDM.

In the second stage, we select eight important items from WLDM's model, which could be measured, and 1088 web sites have been analyzed by considering them and the results of them have been cauterized.

The outcome of our research can be useful for designing the desired web sites, which saves time and price during the design process, our results can be used to propose an efficient and effective web designing to attract more visitors. From the subcategories of three primary components in the Fig. 1 we have used eight features which are more measureable, such as: Cognitive load, Type of web interface, Complexity, The quantity of the links, The quantity of the Images, Web site goal, Font variety and Color variety. The methodology and results help identify user interface issues that a designer should emphasize. Also from the second part of this study, we considered which one of the mentioned results have been used for designing by the world famous web sites..

We provided these data in order to use in third part, which we did not write in this paper and the basic mechanism behind collaborative filtering systems for future work of this research is in the following: A large group of owners' preferences are registered. Using a similarity metric, a subgroup of owners is selected whose preferences are similar to the preferences of the person who has same attitude. Average of the preferences for that subgroup is calculated. We can build a model by data classification and could map the owner group to web sites clusters. In addition, the final software recommends the layout design to new users. In our future work, the model will be used for recommender systems that allow to replace web site designers.

The system with scrutiny of owner's behavior and with the use of psychological methods, suggest him dedicated desired patterns. One of the methods that can be used in this system is investigating the owner's behavior and compares it with another owner using these components. The final aim is making a recommender system as an automatic web designer.

Acknowledgments

We thank the Iran Telecommunication Research Center for financial support.

References

1. Zviran, M., Glezer, C., Avni, I.: User Satisfaction from Commercial Web Sites: The Effect of Design and Use. Information and Management 43(2), 157–178 (2005)
2. Wikipedia (October 2010), http://en.wikipedia.org/wiki/Web_design
3. Parons, J.J., Oja, D.: Computer Concepts, Course Technology, Cambridge, MA (1998)
4. Kazienko, P., Adamski, M.: AdROSA—Adaptive personalization of web advertising. Information Sciences Wrocław, Poland 177(11), 2269–2295 (2007)
5. Garrigós, I., Meliá, S., Casteleyn, S.: Personalizing the Interface in Rich Internet Applications. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 365–378. Springer, Heidelberg (2009)
6. Seo, J., Lee, K.: Development of Web site Design Personalization Service Using Design Recommender System (2003), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.108.3769 (access time: October 2009)
7. Shapira, B., Taieb-Maimon, M., et al.: Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In: ACM Conference (2006)

8. Heimlich, J.E., Wang, K.: Evaluating the structure of web sites. EETAP Resources Library Publications 2 (January 2002)
9. Srivihok, A., Sukonmanee, P.: Intelligent Agent for e-Tourism: Personalization Travel. In: Proc. of The 14 th International World Wide Web Conference Workshop (WWW), pp. 10–14. Keio University, Chiba (2005)
10. Nielsen, J., Tahir, M.: First page Usability 50 Web sites Deconstructed, New Riders, Indianapolis (2002) ISBN
11. Evett, Brown, Evett, L., Brown, D.J.: Text formats and web design for visually impaired and dyslexic readers: clear text for all. The Journal of Interacting with Computers 17, 453–472 (2005)
12. Tullis, T.S.: Screen design. In: Helander, M., Landauer, T.K., Prabhu, P. (eds.) Handbook of Human–Computer Interaction, pp. 503–531. Elsevier Science, Oxford (1997)
13. Scott, D., Findlay, J.M.: The shape of VDU's to come: a visual search study. In: Lovesey, E.J. (ed.) Contemporary Ergonomics, pp. 353–358. Taylor & Francis, London (1990)
14. Foltz, P.W., Kintsch, W., Landauer, T.K.: The measurement of textualcoherence with Latent Semantic Analysis. Discourse Processes 25, 285–307 (1998)
15. Norman, K.L.: Better design of menu selection systems through cognitive psychology and human factors. Human Factors (Golden Anniversary Special Issue) 50, 556–559 (2008)
16. Tracy, J.P., Albers, M.J.: Measuring Cognitive Load to Test the Usability of Web Sites. In: Proceedings of the Annual Conference for Technical Communication, pp. 256–260 (2006)
17. Schiaffino, S., Amandi, A.: User–interface agent interaction: personalization issues. International Journal of Human-Computer Studies 60(1), 129–148 (2004)
18. Ignacio Madrid, R., Van Oostendorp, H., Puerta Melguizo, M.C.: The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. Computers in Human Behavior 25(1), 66–75 (2009)
19. Nygren, E.: From paper to computer screen: Human information-processing and user interface (1996)
20. Chalmers, P.: The role of cognitive theory in human-computer interface. Computers in Human Behavior 19, 593–607 (2003)
21. Ling, J., van Schaik, P.: The influence of font type and line length on visual search and information retrieval in web pages. Int. J. Hum.-Comput. Stud. 64(5), 395–404 (2006)
22. Micheal, B., Lida, B., Riley, S., Hackler, T., Janzen, K.: A Comparison of Popular Online Fonts: Which Size and Type is Best? Usability News 4(1) (January 2002)
23. Weller, D.: The Effects of Contrast and Density on Visual Web Search, `http://psychology.wichita.edu/newsurl/usabilitynews/62/density.asp`
24. Welling, L., Thomson, L.: PHP and MySQL Web Development, 3rd edn. Sams Publishing, USA (2005)
25. SuperPower: Visualising the internet (January 2010), `http://news.bbc.co.uk/nol/shared/spl/hi/sci_nat/10/internet_treemap/assets/nielsen_top_100_january_2010b.xls`
26. The 1000 most-visited sites on the web (December 2010), `http://www.google.com/adplanner/static/top1000/`
27. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
28. Aittokallio, T., Malminen, J., Pahikkala, T., Polo, O., Nevalainen, O.: Inspiratory flow shape clustering: An automated method to monitor upper airway performance during sleep. Comput. Methods Progr. Biomed. 85, 8 (2007)

29. Cho, Y., Kim, J., et al.: A personalized recommender system based on web usage mining and decision tree induction. Expert Systems with Applications, World Scientific and Engineering Academy and Society (WSEAS) 23(3), 329–342 (2002)
30. Melville, P., Mooney, R., et al.: Content-boosted collaborative filtering for improved recommendations. AAAI Press; MIT Press, Menlo Park, CA (2002)
31. Huang, K.J., et al.: A survey of E-commerce recommender systems. Service Systems and Service Management. In: International Conference on In Service Systems and Service Management, pp. 1–5 (2007)
32. Shapira, B., Taieb-Maimon, M., et al.: Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In: ACM Conference (2006)
33. Stormer, H.: Personalized web sites for mobile devices using dynamic cascading style sheets. International Journal of Web Information Systems 1(2), 83–88 (2005)
34. Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum, Hillsdale (1983)
35. Sheppard, S., Rouff, C.: In: John Marciniak, J. (ed.) Encyclopedia of Software Engineering, vol. 2, John Wiley & Sons, New York (1994)
36. Kintsch, W.: The role of knowledge in discourse comprehension: A construction integration model. Psychological Review 95, 163–182 (1988)
37. Templatemonster (October 2010), `http://www.templatemonster.com`
38. Nahm, E.S., Preece, J., Resnick, B., Mills, M.E.: Usability of health web sites for older adults: A preliminary study. CIN: Computers, Informatics 22, 326–334 (2004)
39. Kolers, P.A., Duchnicky, R.L., Ferguson, D.C.: Eye movement measurement of readability of CRT displays. Human Factors 23, 517–527 (1981)
40. Bernard, M., Liao, C.H., Mills, M.: The effects of font type and size on the legibility and reading time of online text by older adults. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, CHI 2001, Seattle, Washington, March 31 - April 05, pp. 175–176. ACM Press, New York (2001)

Interacting with Semantics and Time

Christian Stab, Kawa Nazemi, Matthias Breyer, Dirk Burkhardt, and Arjan Kuijper

Fraunhofer Institute for Computer Graphics Research,
Fraunhoferstrasse 5, 64283 Darmstadt, Germany
{christian.stab,kawa.nazemi,matthias.breyer,dirk.burkhardt,
arjan.kuijper}@igd.fraunhofer.de

Abstract. Time appears in many different semantic information systems like historical databases, multimedia systems or social communities as a common attribute. Beside the temporal information, the resources in these domains are categorized in a domain-specific schema and interconnected by semantic relations. Nevertheless, the high potential of these systems is not yet exhausted completely. Even today most of these knowledge systems present time-dependent semantic knowledge in textual form, what makes it difficult for the average user to understand temporal structures and dependencies. For bridging this gap between human and computer and for simplifying the exploration of time-dependent semantic knowledge, we developed a new interactive timeline visualization called SemaTime. The new designed temporal navigation concept offers an intuitive way for exploring and filtering time-depended resources. Additionally SemaTime offers navigation and visual filtering methods on the conceptual layer of the domain and is able to depict semantic relations. In this paper we describe the conceptual design of SemaTime and illustrate its application potentials in semantic search environments.

Keywords: Information Visualization, Semantic Visualization, Timeline Visualization, Semantic Search, Time-Dependent Ontology.

1 Introduction

In many different information systems time appears as a common attribute. For example the resources of historical databases contain timestamps for describing the chronological order of events [1], media files in multimedia systems are tagged with release dates [2] or profiles in social communities contain different time-dependent information about a certain person [3,4]. Additionally the resources in these systems are often specified by using semantic technologies which offer new possibilities for modeling and representing information in modern knowledge systems. Beside the categorization in a domain-specific schema, semantic description languages allow the modeling of relationships between different resources [5] and offer new potential for improving search mechanisms, navigation patterns and graphical user interfaces.

Nevertheless, the high potential of these information systems is not yet exhausted completely, in particular when it comes to imparting the contained information to the user. Even today most of these systems present time-dependent and semantic

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 520–529, 2011.

information in textual form, what makes it difficult for the average user to understand the temporal structure and dependencies between the resources. Graphical User Interfaces (GUI) and Information Visualizations (IV) can bridge this gap between human and computers and simplify the handling of complex time-dependent and semantic knowledge [6].

In this paper we present a novel interactive timeline visualization called SemaTime that is especially designed to present time-dependent semantic information to the user. This visualization is based on the well-known timeline metaphor that arranges time-dependent entities along a time-axis. The new designed temporal navigation concept offers an intuitive way for exploring and filtering temporal resources as well as simplified navigation mechanisms for the temporal dimension. The dual time-axis concept supports the user to find the demanded information faster and offers an orientation on the temporal selection. In order to prevent context-loss, an overview is integrated in the time-axis for visualizing context information of the currently selected time interval.

The remaining paper is structured as followed: In the next section we examine existing timeline visualizations and discuss their ability to visualize time-dependent semantic knowledge. Afterward we present the conceptual design of SemaTime and give a detailed description of its interaction and navigation mechanisms. We conclude this paper with an application scenario, where we illustrate the application potential of SemaTime in a semantic search environment.

2 Related Work

Nowadays there are many different approaches for visualizing time-dependent information (e.g. [7,8]). Most of these visualizations are based on timeline visualizations that arrange time-dependent entities along a time-axis. For example Allen [9] uses an interactive timeline for the chronological visualization of the content of a digital library. This timeline visualizes past events and periods. To indicate different categories, the visualized entities can be illustrated with different colors but the visualization of hierarchical categories or semantic relations is not supported by this approach. A widely used timeline visualization is SIMILE [10] and is used from Alonso et al. [11] to visualize search results. SIMILE provides the presentation of events and periods along a horizontal time axis and offers different navigation features and an overview function. The Context-Focus-Timeline [1] visualizes events on a vertical time-axis and is used in a History Event Browser. For every event further information can be displayed whereas the timeline offers a clear overview of the temporal context.

However for the visualization of time-dependent semantics the visualization should be able to present the domain-specific, hierarchical schema and semantic relations between the entities. One approach for visualizing these complex data structures is introduced by Plaisant et al. [12]. Lifelines is able to visualize different facets which are depicted as horizontal slices and optionally hierarchical structured. Lifelines even supports the visualization of semantic relations in an implicit way by highlighting related events or periods on demand. However there is no possibility for distinguishing different relation types or to understand the meaning of relations.

Bui et al. introduce an interactive timeline visualization for depicting medical patient records in hospital information systems [13]. This timeline visualization uses a similar approach for visualizing hierarchical categories but is not able to visualize semantic relations. Another timeline visualization that supports the pictorial representation of hierarchical structured and time-dependent information is Timeline Trees [14] which is designed for visualizing sequences of transactions in information hierarchies. The hierarchical structure of the available data is depicted as horizontally oriented tree visualization. The nodes of the tree can be expanded or collapsed to support information filtering and to navigate through the hierarchical structure.

One of the first approaches for explicitly depicting semantic relations in timeline visualizations is introduced by Kumar et al. [15]. The tmVIEWER is able to visualize beside events and periods, relations as directed edges between time-dependent entities. However, this approach can only conditionally be applied to visualize semantic relations in a temporal domain because semantic relations may change in the course of time or may have a certain period of validity (e.g. “works_at”, “lives_in”, etc.). Jensen uses the same approach for depicting relations in the SemTime visualization [16]. SemTime also provides stacking of timelines and supports the visualization of flat categories in different layers which can be independently set to different time intervals.

Beside the introduced timeline visualizations there are several other approaches that are especially designed for a particular application or a specific kind of time-dependent information. For example the timeline visualization from Bade et al. [17] is especially designed for visualizing high-dimensional, time-oriented data and is used in the area of intensive care units e.g. for depicting fever curves of patients. André et al. introduce the timeline visualization Continuum [18] that comes with user-determined controls over the level of detail, a histogram function and a comparative split view. Furthermore Continuum is able to visualize time-dependent, hierarchically structured entities by nesting. However this approach is only suitable for visualizing hierarchies in which every node is time-dependent and for this reason it is not applicable for visualizing a general, hierarchical categorization.

3 The SemaTime User Interface

SemaTime is a visualization component of the SemaVis-Framework[1] [19], an adaptive visualization framework that contains different aspect-oriented visualizations [20, 21] for semantic knowledge. All components and visualizations are implemented in Adobe Flex[2] and can be connected with heterogeneous semantic data bases (for example via SPARQL). SemaVis also offers different functionalities for combining several visualization components to one knowledge cockpit [22] that offers different aspect oriented perspectives of the same information. In the following, we first explain the different parts of SemaTime and their basic functionality before describing the mechanisms in more detail.

---

[1] http://www.Semavis.com

[2] http://www.Adobe.com/products/flex

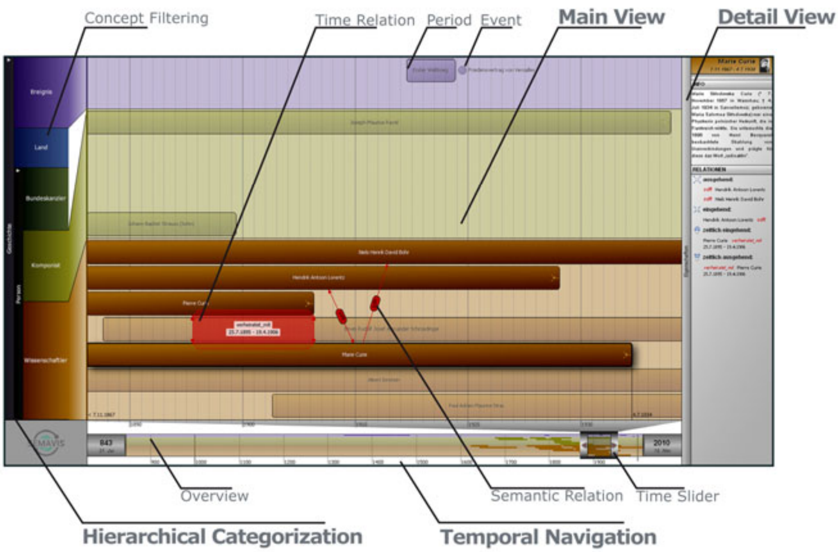


Fig. 1. The user interface of SemaTime is divided into four parts: Temporal Navigation (bottom), Hierarchical Categorization (left), Main View (center) and Detail View (right)

The user interface of SemaTime is designed like a common two-dimensional coordinate system (Fig. 1) in which the x-axis marks the temporal dimension and the y-axis represents the hierarchical categorization of the given domain. Figure 1 shows a screenshot of the SemaTime User Interface that is divided into the following four parts:

- The *Temporal Navigation Panel* at the bottom of the user interface is divided into two timescales. It offers an overview visualization and different methods for navigating in the temporal dimension.
- The *Hierarchical Categorization* of the domain is depicted as the y-axis on the left side of SemaTime. It divides the Main View in different horizontal slices and offers navigation and visual filtering methods on the conceptual layer of the domain.
- The *Main View* in the center of the user interface arranges time-dependent entities and relations between them according to their temporal information and corresponding category.
- The *Detail View* on the right side of the visualization offers further information about the selected entity e.g. descriptions, images or links to external sources.

3.1 Temporal Navigation

One of the essential parts of any visualization for time-dependent information is an appropriate navigation and filtering mechanism for the temporal dimension. On the one hand the user should be able to navigate through the temporal dimension and to explore time-dependent entities in a certain time interval of interest. On the other

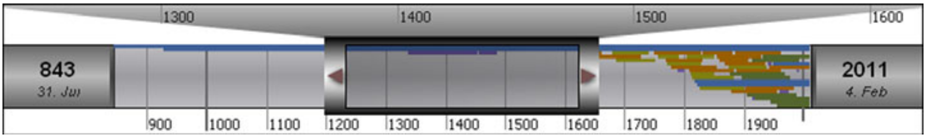


Fig. 2. The Temporal Navigation Panel in SemaTime offers two time axes for the navigation in the temporal dimension and overview visualization

hand an appropriate navigation function should assist the user in keeping the overview about the temporal information and offer an orientation on the current temporal selection.

The Temporal Navigation Panel at the bottom of SemaTime (Fig. 2) is divided into two timescales: (1) A global timescale over the complete time interval of the given data and (2) a selection scale over the currently selected time interval. The global scale contains a slider that allows the selection of a certain time interval and thus an adjustment of the selection scale. Additionally, the global time scale contains an overview visualization. So users can be aware of entities that are not visible in the Main View with the currently selected time interval.

For navigating in the temporal dimension and for selecting a certain time interval, SemaTime offers the following interactions: (1) A temporal zoom and (2) a temporal pan navigation. The temporal zoom allows the user to zoom in or out of the currently selected time interval by using the mouse wheel. So it is possible, to adjust the granularity regarding the temporal dimension and to view the information e.g. in centuries, decades or month. Since the temporal zoom influences both boundaries of the selected time interval, it is additionally possible to adjust the boundaries individually by dragging the left or right side of the time slider on the global scale. Thus it is easier to select a specific time interval and the navigation in the temporal dimension is facilitated for the users.

In contrast to the temporal zoom the pan navigation implemented in SemaTime allows the temporal navigation without influencing the temporal granularity. For this type of navigation there are two different interactions available. On the one hand the user is able to pan the selected area by dragging the time slider on the global scale either to the left or to the right side. The other option is to click and drag directly in the Main View.

3.2 Navigation and Filtering on the Conceptual Layer

Semantically modeled resources are categorized in a hierarchical domain-specific schema. This conceptual layer of the semantic structure divides the available information into knowledge spaces that cover a particular subject. SemaTime visualizes this categorization schema in the Hierarchical Categorization View (Fig. 3) to offer beside the temporal navigation, visual filtering and navigation mechanisms on the conceptual layer. So the user is able to navigate from an abstract level through the semantic knowledge space and locate specific resources of interest. This component of SemaTime divides the Main View into horizontal slices each of which corresponds

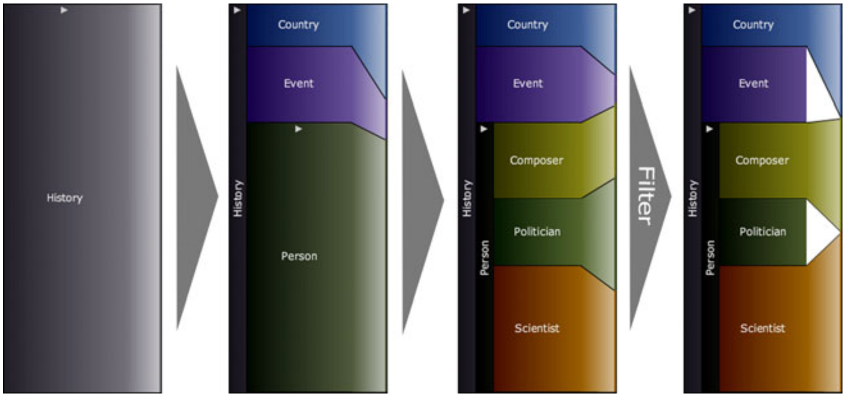


Fig. 3. The Hierarchical Categorization visualizes the domain-specific schema and offers navigation and filtering functions on the conceptual layer

with a category in the hierarchy. By clicking on a node it is possible to expand or collapse the visualized categories and to filter specific nodes.

The sizes of the displayed categories are calculated relative to the amount of resources in a category. For instance the dataset in Figure 3 contains more persons than countries. Due to the layout algorithm of the Main View that calculates the minimal number of lines and displays every resource with the same height, the heights of categories in the Hierarchical Categorization View and the Main View are not identical. For this reason, a trapezoid view for connecting the Hierarchical Categorization View with the Main View is integrated to ensure ideal space filling and visual concept assignment in the Main View even if elements are filtered. To increase the user experience with SemaTime, the transitions during navigation or filtering procedures are smoothly animated.

3.3 Visualizing Time-Dependent Resources

The Main View in the center of SemaTime visualizes time-dependent resources and places them according to their temporal attributes and associated category. SemaTime differentiates between two different types of time-dependent resources: (1) Events and (2) periods. In contrast to periods, events are characterized by a single timestamp. They correspond to a specific date on which the event occurred. On the other hand, time periods are characterized by a unique starting and endpoint, whereby a certain time interval is defined. In order to assure a clear visibility for each temporal granularity and a visible differentiation of these two types, events are visualized as a circle and time periods are depicted as rectangles (Fig. 4). This representation has the advantage that events are still visible at large selected time intervals and are not visualized as thin lines. Figure 4 shows an example of an event (Treaty of Versailles) and a period (World War I) depicted in SemaTime. Since the exact time of a resource may be difficult to recognize in cases of large selected time intervals in the Main View, the user is able to display this information on demand. Therefore, the user

Fig. 4. Visualization of events and periods in SemaTime

selects an element by clicking or a mouse hover action whereby further information becomes visible (Fig. 4).

3.4 Visualization of Semantic Relations in a Temporal Environment

Semantic relations constitute the centerpiece of a semantically modeled domain. They are used to model links between resources and provide additional information about knowledge entities. In particular in time-dependent domains, semantic relations can be used to model causal dependencies. Commonly, a relation between two resources is defined by a direction and a type and is depicted by interconnecting the resources with a directed and labeled edge. However in a temporal domain it is also possible that a semantic relation is only valid for a certain time interval or changes in the course of time. Hence a semantic relation may contain temporal information that should be adequately represented. For this reason SemaTime distinguishes between time-dependent relations and static relations without time reference (Fig. 5).

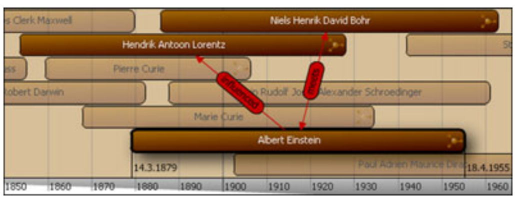


Fig. 5. Semantic relations without time reference

Time-dependent relations between two resources are depicted as a rectangle that denotes the temporal validity of the relation. The direction of the relation is visualized by arrows pointing to the related entity. The type of the relation and the temporal information are visualized in a label inside of the rectangle. Figure 6 shows an example of a bilateral, time-dependent relation. It is easy to recognize that Marie Curie was married with Pierre until his death in 1906 and vice versa.

Depending on the modeled domain and the application, the given semantic structure may contain a huge number of interconnections. The visualization of this structure may result in overlapping edges in particular in a temporal visualization in which the placement of resources is defined by temporal properties. For this reason, semantic relations in SemaTime will only be displayed on demand. Additionally, all semantically not related resources are reduced in their alpha values if a resource is selected with a mouse click. Thus the semantic surrounding of the selected resource is highlighted without losing the context in the temporal environment.

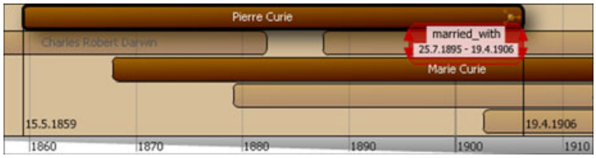


Fig. 6. Time-dependent relation in SemaTime

4 Visualization of Search Results in Semantic Domains

SemaTime was developed as a visualization component of the SemaVis Framework [19], a development of the Core-Technology-Cluster (CTC) Semantics Visualization and Innovative User Interfaces of THESEUS [23]. THESEUS is a 60-month program partially funded by the German Federal Ministry of Economics and Technology. The SemaVis Framework provides core technologies for visualization, editing and annotation of semantically annotated data. The main goal of SemaVis is the provision of core-technologies for heterogeneous users, data and application scenarios.

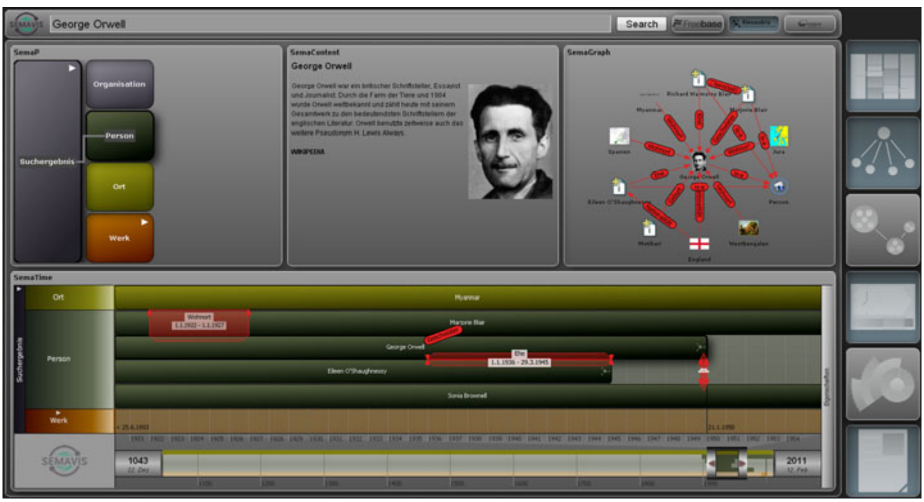


Fig.7. SemaTime in a semantic search scenario

Figure 7 shows SemaTime in an orchestrated knowledge cockpit [22] with other visualizations of the SemaVis-Framework (e.g. SeMap [21]) in a semantic search scenario. In this example the semantic knowledge base of the THESEUS Use Case ALEXANDRIA provided by neofonie[3] is searched for the author George Orwell and the results are visualized by the SemaVis technologies. In this scenario SemaTime offers beside the birth and death date, information about certain stages in his life. A time-dependent relation in this example depicts that he lived for a period of five years

[3] http://www.Neofonie.de

in Myanmar/Burma. Two other temporal relations present both of his marriages. And it is recognizable that he was married with Eileen O'Shaughnessy until her death in 1945. Additionally, to the demonstrated potential and benefit of visualizing search results with SemaTime, further information e.g. relations to non-temporal resources, a short description and images are illustrated with other visualizations of SemaVis.

5 Conclusion and Outlook

In this paper we introduced a novel interactive timeline visualization that visualizes the temporal structure and relations between resources of time-dependent semantic domains. The user interface of SemaTime includes a novel temporal navigation concept that offers an intuitive way for exploring and filtering temporal resources. Furthermore, the introduced visualization uses the domain-specific schema to offer visual filtering and navigation mechanisms on the conceptual layer. SemaTime can therefore save a lot of time during the search of temporal relationships and structures.

The further progress of this work includes a comprehensive evaluation that further examines the potentials and benefits of visualizing time-dependent semantics. In particular the future work includes a comparison with common presentations with respect to the discovery and analysis of time-dependent, semantically modeled resources.

References

1 Allen, R.B.: A focus-context browser for multiple timelines. In: JCDL 2005: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 260–261. ACM, New York (2005)
2 Ames, M.G., Manguy, L.: PhotoArcs: ludic tools for sharing photograph. In: MULTIMEDIA 2006: Proceedings of the 14th annual ACM international conference on Multimedia, pp. 615–618. ACM, New York (2006)
3 Crow, J., Whitworth, E., Wongsa, A., Francisco-Revilla, L., Pendyala, S.: Timeline interactive multimedia experience (time): on location access to aggregate event information. In: JCDL 2010: Proceedings of the 10th Annual Joint Conference on Digital Libraries, pp. 201–204. ACM, New York (2010)
4 Erickson, T., Laff, M.R.: The design of the ’Babble’ timeline: a social proxy for visualizing group activity over time. In: CHI 2001 extended abstracts on Human Factors in Computing Systems, pp. 329–330. ACM, New York (2001)
5 Studer, R., Davies, J., Warren, P.: Semantic Web Technologies: Trends and Research in Ontology-Based Systems. Wiley, Chichester (2006)
6 Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology Visualization Methods - A Survey. ACM Comput. Surv. 39, 10:1–10:43 (2007)
7 Aigner, W., Miksch, S., Müller, W., Schumann, H., Tominski, C.: Visualizing time oriented data-a systematic view. Comput. Graph. 31, 401–409 (2007)
8 Silva, S.F., Catarci, T.: Visualization of linear time-oriented data: A survey. In: WISE 2000: Proceedings of the First International Conference on Web Information Systems Engineering (WISE 2000), vol. 1, pp. 310–319. IEEE Computer Society, Los Alamitos (2000)

9 Allen, R.B.: Interactive timelines as information system interfaces. In: Symposium on Digital Libraries, Japan, pp. 175–180 (1995)
10 SIMILE Widgets - Timeline Web Widget for Visualizing Temporal Data (2011), `http://www.simile-widgets.org/timeline`
11 Alonso, O., Baeza-Yates, R., Gertz, M.: Exploratory search using timelines. In: SIGCHI 2007 Workshop on Exploratory Search and HCI Workshop (2007)
12 Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B.: Lifelines: visualizing personal histories. In: CHI 1996: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 221–227. ACM, New York (1996)
13 Bui, A.A.T., Aberle, D.R., Kangarloo, H.: Timeline: Visualizing integrated patient records. IEEE Transactions on Information Technology in Biomedicine 11, 462–473 (2007)
14 Burch, M., Beck, F., Diehl, S.: Timeline trees: visualizing sequences of transactions in information hierarchies. In: AVI 2008: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 75–82. ACM, New York (2008)
15 Kumar, V., Furuta, R., Allen, R.B.: Metadata visualization for digital libraries: interactive timeline editing and review. In: DL 1998: Proceedings of the Third ACM Conference on Digital Libraries, pp. 126–133. ACM, New York (1998)
16 Jensen, M.: Visualizing complex semantic timelines. NewsBlib (2003)
17 Bade, R., Schlechtweg, S., Miksch, S.: Connecting time-oriented data and information to a coherent interactive visualization. In: CHI 2004: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 105–112. ACM, New York (2004)
18 André, P., Wilson, M.L., Russell, A., Smith, D.A., Owens, A., Schraefel, M.C.: Continuum: designing timelines for hierarchies, relationships and scale. In: UIST 2007: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, pp. 101–110. ACM, New York (2007)
19 Nazemi, K., Breyer, M., Burkhardt, D., Stab, C., Hofmann, C., Hornung, C.: D.CTC.5.7.1: Design und conceptualization semantic visualization framework. THESEUS Deliverables: CTC-WP5 Innovative User Interfaces and Visualization 2 (2009)
20 Stab, C., Breyer, M., Nazemi, K., Burkhardt, D., Hofmann, C., Fellner, D.W.: SemaSun: Visualization of Semantic Knowledge based on an improved Sunburst Visualization Metaphor. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, vol. 2010, pp. 911–919. AACE, Chesapeake (2010)
21 Nazemi, K., Breyer, M., Hornung, C.: SeMap: A Concept for the Visualization of Semantics as Maps. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 83–91. Springer, Heidelberg (2009)
22 Nazemi, K., Breyer, M., Burkhardt, D., Fellner, D.W.: Visualization Cockpit: Orchestration of Multiple Visualizations for Knowledge-Exploration. International Journal of Advanced Corporate Learning (iJAC) 3(4) (2010)
23 Theseus - New Technologies for the Internet of Services (2011), `http://theseus-programm.de`

Investigating Drag and Drop Techniques for Older People with Cognitive Impairment

Frédéric Vella[1], Nadine Vigouroux[1], and Pierre Rumeau[2]

[1] IRIT, Université Paul Sabatier, 118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9
[2] Laboratoire de Gérontechnologie La Grave CHU Toulouse / Inserm U 558 / Gérontopôle, 37 Allées Jules Guesde, 31073 Toulouse Cedex
{vella,vigourou}@irit.fr, rumeau.p@chu-toulouse.fr

Abstract. Graphical user interfaces and interactions that involve pointing to items and dragging them are becoming more common in rehabilitation and assistive technologies. We are currently investigating interaction techniques to understand point-select-drag interactions for older people with cognitive impairment. In particular, this study reports how older perform such tasks. Significant differences in behavior between all of the interaction techniques are observed and the reasons for these differences are discussed according the Mini Mental Score.

Keywords: Human Computer Interaction, Older subjects, dementia, evaluation.

1 Introduction

Older people with cognitive impairment are a challenging user group to study the human computer interaction. Because of the increased invasion of computer and portable devices in world and the demographic "ageing" of people, there is a need for usable interface that promotes accessibility and usability for older people. Another effect of the demographic shift towards an ageing population is a growing number of people with dementia. Current estimates put the numbers of people with dementia at around 25 million worldwide but this is predicted to rise to 114 millions [16].

The majority of people with dementia are over 65, and most likely have Alzheimer's disease, cerebrovascular disease or a combination of both [14].

Graphical user interfaces and interactions that involve pointing to items and dragging them are becoming more common in rehabilitation and assistive technologies. It is why there is an important need to study computer pointing devices for older people with cognitive impairment.

While many studies have examined age-related declines in motor control trough experimental observations showing that older people have slower movement times and greater difficulty producing fine motor adjustments [2]. Many studies have also looked at the effects of ageing on computer interaction [1], [3], [4], [5], Chaparro & al [5] have investigated age-related differences using these two devices (e.g. mouse and trackball). The results showed that older people moved more slowly than younger adults by two groups of adults (younger < 40 years of age and older > 65 years older).

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 530–538, 2011.

Another point is that no age differences were found in movement time or variable error between the two devices. Smith & al. [4] have also examined age differences in the performance of basic computer mouse device. Differences in performance due to age were found for complex tasks (clicking and double-clicking).

Many approaches in the mainstream human-computer interaction community have been proposed to ease pointing tasks on computers, although few have been directed at older adults [e.g. 3, 6, 7, 8, 9 & 10]. However, these essentially aim to improve point and click interaction, using familiar pointing devices, such as the computer mouse, which manipulate an on-screen pointer.

One of the key challenges in developing computer accessibility is finding some easy and effective means of interaction. This is even more necessary in the process of designing interactive systems that could rehabilitate them such serious games, memory aids, history books, etc.

A major point from reported above is that older subjects take longer to realize selections, and require a greater proportion of time and a higher number of corrective movements to reach the targets with the same level of accuracy as younger subjects. So, compared with younger users, older subjects can have greater difficulty to perform the aiming, clicking, and movements required to point-select-drag interactions.

Therefore, the goal of this study is to investigate how older people with cognitive impairment respond to different drag and drop interaction techniques. This ongoing study will present our experiment and Mini Mental State group differences in movement time, movement distance and mean number of mistakes.

2 Experiment

This experiment studies three interaction techniques to select and to move an item to another one with a mouse. The aims of this experiment are to:

- Analyze performance differences between older subjects without cognitive impairment and with cognitive impairment;
- Identify and analyze factors that can explain any preference or better efficiency observed between these interaction classes.

2.1 Materials

The experiments were conducted on a Satellite Pro A200Toshiba laptop with a 15 inch widescreen, 1024*768 TFT display. An optical computer mouse was used as an input device. The right button was deactivated. The mouse was selected to be representative of an input device that would be typical for personal use. A hand cursor was preferred as a good metaphor.

2.2 Tasks

Study participants (Fig.2) have been asked to perform some serial pointing and moving tasks (Fig.1). They have been asked to select the item (the piece of sugar) and to put it in the middle of the cup of coffee. The size of sugar is 57 x 38 pixels. The distance between the sugar center and one cup of coffee is 406 pixels. Three modes of the “drag and drop” interaction techniques have been defined:

Fig. 1. Exercise view

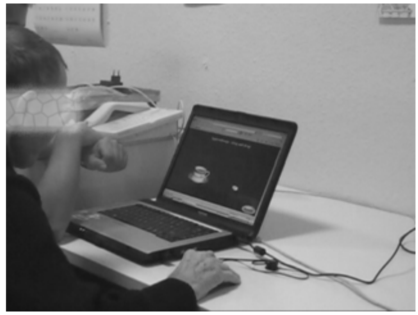


Fig. 2. A subject doing the exercise

- **Clicking Interaction** (CL)**:** The subject selects the sugar by clicking it, moves the cursor to the coffee cup top, and clicks in the area of the coffeethen the piece of sugar is falling down;
- **Dragging Interaction** (DR)**:** The subject selects the sugar by clicking it, maintains the pressure on the mouse button until the cursor is over the cup of coffee, then release the button and the sugar is falling in;
- **Clicking and Magnetization Interaction** (CAM)**:** The subject selects the sugar by clicking it, then the sugar is automatically attached to the cursor; he moves the cursor over the cup of the coffee, and the sugar is falling in automatically.

For this experiment, the size of the sugar target and the distance between the sugar and the cup of coffee is fixed. A sound feedback is playing to inform that the sugar is taken by "hands" during the clicking action. A splash sound is playing when the piece of the sugar is falling down in over the coffee.

2.3 Experimental Design

The study utilized a split design with task (CL, DR and CAM).

2.4 Procedure

Participants sat approximately 60 cm from the screen with the mouse positioned for right-hand use as default configuration. Participant sessions involved a set of training and trial computer sessions followed by a semi-structured questionnaire.

The training phase consisted in: firstly, describing the run of the mouse (moving and clicking principles) to encourage familiarity with the device, secondly doing the exercise with each "drag and drop" interaction technique. We have considered that the technique was mastered when the subject was capable of using it without any comment or help from the experimenter. This was followed by the experimental session. On the display, participants viewed a sugar piece and a cup of coffee. The task involved moving the cursor to the sugar piece and then to the cup of coffee according the "drag and drop interaction". Each participant performed a total of nine trials (3 trials per each task). Tasks were counterbalanced.

The session trial is considered succeeded when the nine trials were The questionnaire was designed to complement the movement behavior and to address issues such as computer expertise, preferred interaction technique, difficulties of computer use, etc.

2.5 Participants

Physicians' geriatrics division of the Center Hospitalize Universities de Toulouse recruited 97 subjects, aged over 65 years with and without Alzheimer's disease. All these subjects have accepted to participate to the experiment. The Mini Mental State (MMS) examination was made by an geriatric doctor in Alzheimer disease.

They were classified into 5 groups according to their cognitive impairment dementia estimated on the MMS [12] We have used the Feldmann and Woodward's distribution [11] which consists into five intervals.

Subjects cannot understand the instructions or cannot communicate, have not been recruited by doctors

Table 1. Participants according to MMS and success

MMS	**<10**	**[10-14]**	**[15-20]**	**[21-26]**	**[27-30]**
Number of subjects	8	16	19	26	28
Failure (F) and Achievement (A)	7F/1A	7F /9A	9F/10A	3F/23A	1F /27A

The table 1 confirms more the MMS is low more the subjects were forced to withdraw the exercise (7/16 in the [10-14] and 78 in the [0-9] intervals.

3 Results

In this study, we have not considered the class of MMS <10 because one subject was successful exercises. The results were not significant for this class.

The time, the distance as well the mistakes were analyzed for differences between MMS intervals and techniques interaction. The figures in this section show mean values for each group and interaction techniques, with bars showing the standard deviation.

3.1 Average Time for the 3 Interactions Techniques

This parameter is the time to move the sugar in the coffee. One major significant (Fig.3) result is that the duration factor is significantly different for the three interaction techniques: the DR duration is much longer than this CAM and CL (CL=16 s, DR=19 s and CAM=6 s). This result is independent of the age.

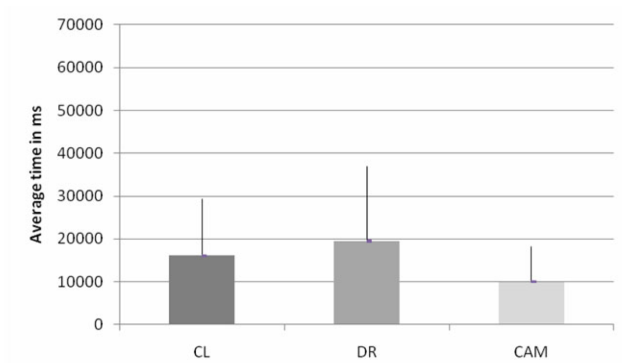


Fig. 3. Average time for using of the 3 interactions techniques

Another important result is that the duration increase is correlated to the decrease of the Mini Mental Score (MMS) for CL and DR (Fig.4). We observe also large behaviour variability for all MMS class.

CAM (Fig.4) duration is not also dependant of the MMS. From the empirical observations, we identify several difficulties with mouse, such as losing the cursor and bad control in moving with DR.

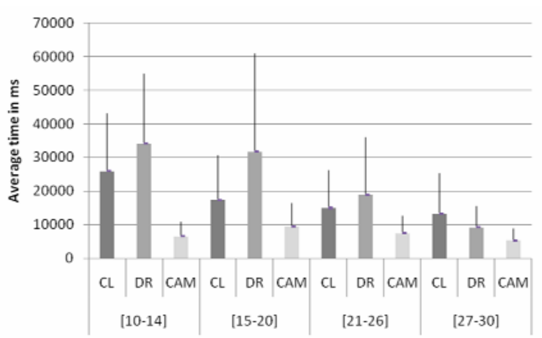


Fig. 4. Average time for using of the 3 interactions techniques according to MMS

3.2 Average Distance of Cursor for the 3 Interactions Techniques

This parameter is the distance to move the sugar in the coffee. This result (Fig.5) showed this distance is more important for DR and CL than CAM.

The Fig.6 showed that the distance increase is correlated to the decrease of the Mini Mental State (MMS) for CL and DR. We observe also large behavior variability for all MMS class. CAM distance does not vary according to the MMS.

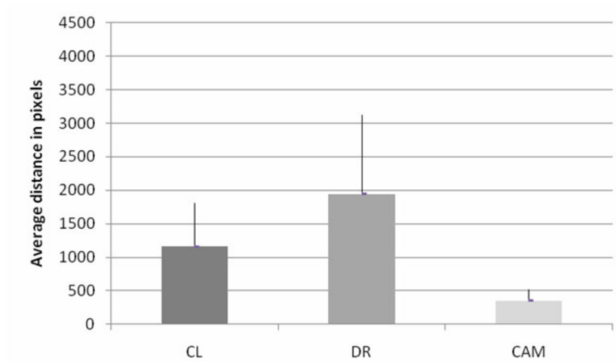


Fig. 5. Average distance of cursor for the 3 interactions techniques

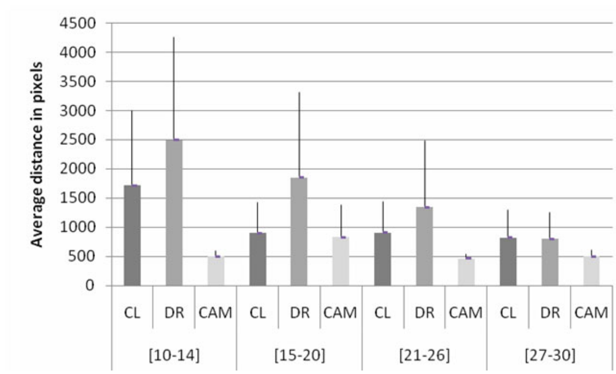


Fig. 6. Average distance of cursor for the 3 interactions techniques according to MMS

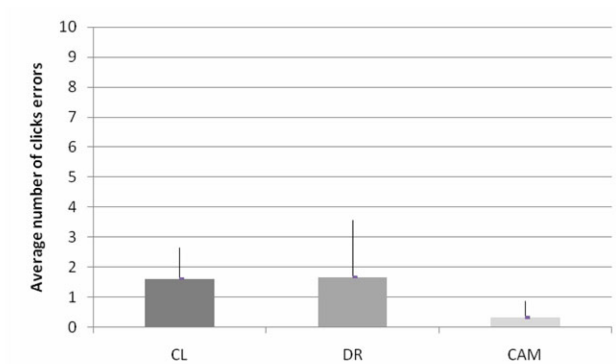


Fig. 7. Average number of clicks errors for the 3 interactions techniques

3.3 Average Number of Clicks Errors for the 3 Interactions Techniques

The numbers of clicks errors are the supplementary clicks during the move of the sugar in the coffee. This result (Fig.7) showed these clicks errors are more important for DR and CL than CAM.

The Fig.8 showed that the clicks errors increase is correlated to the decrease of the Mini Mental Score (MMS) for CL and DR. We observe also large behavior variability for MMS [10-14] and [15-20]. The clicks errors of CAM do not vary according to the MMS. We observe a small variability.

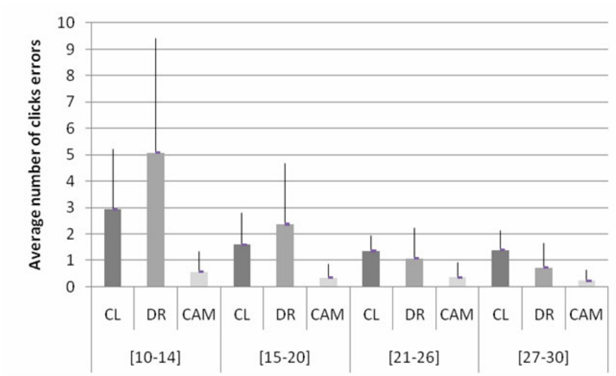


Fig. 8. Average number of clicks errors for the 3 interactions techniques according to MMS

4 Discussions

The CAM interaction is well appreciated because the technique represents well the natural actions (taking and moving). It is a good metaphor. The DR was rejected by subjects because it requests too much workload: this fact can be explained because two simultaneous processes (moving and pressure) are involved in the task.

When the subject was falling in one of the part of exercise (clicking or moving with or without pressure), he/she developed two main behaviours: one is asking help (for instance, can you show me, can you explain me, what do must do now?), another is doing with hand as he/she would have done in a real world.

In the results we note that the MMS has an important role on the use of DR technique and CL technique. Regarding the DR, it requires the user to perform two simultaneous actions moving and clicking. On CL, it has the disadvantage of moving the object automatically. This disturbs users because in reality objects do not move alone. However, we note that MMS varies very little on the CAM technique. In fact, it behaves as an aid to move the object that is hooked to the cursor. We note that, whatever the MMS, the distance travelled (average 570 pixels) the cursor is very close to the shortest distance between sugar and coffee (406 pixels).

5 Conclusion and Future Works

This paper has focused on presenting the experiment on point-select-drag interactions. One of the most important points is that the pointing interaction technique has an important impact on the cognitive activity of the subject. We noted that the CAM technique is the more adapted because the MMS varies little. We plan to model the tasks (move, click) of the older people with cognitive impairment. This model allowed to design and to evaluate IHM for them.

References

1 Walker, N., Millians, J., Worden, A.: Mouse accelerations and performance of older computer users. In: Proceedings of Human Factors and Ergonomics Society, 40 th Annual Meeting, Philidelphia, PA, pp. 151–154 (1996)

2 Walker, N., Philbin, D., Fisk, A.: Age-related differences in movement control: Adjusting submovement structure to optimize performance. Journal of Gerontology, Psychological Sciences 52, 40–52 (1997)

3 Worden, A., Walker, N., Bharat, K., Hudson, S.: Making computers easier for older adults to use: area cursors and sticky icons. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA (1997)

4 Smith, M.W., Sharit, J., Czaja, S.J.: Aging, Motor Control, and the Performance of Computer Mouse Tasks. Human Factors 41(3), 389–396 (1999)

5 Chaparro, A., Bohan, M., Fernandez, J.E., Choi, S.D., Kattel, B.: The impact of age on computer input device use: Psychophysical and physiological measures. International Journal of Industrial Ergonomics 24, 503–513 (1999)

6 Trewin, S., Keates, S., et al.: Developing steady clicks: a method of cursor assistance for people with motor impairments. In: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, ACM Press, Portland (2006)

7 Hourcade, J.P., Nguyen, C.M., Perry, K.B., Denburg, N.: PointAssist for Older Adults: Analyzing Sub-Movement Characteristics to Aid in Pointing Tasks. In: CHI 2010, Seniors Using technologies, Atlanta, GA, USA, April 10-15, pp. 1115–1124 (2010)

8 Wood, E., Willoughby, T., Rushing, A., Bechtel, L., Gilbert, J.: Use of Computers Input Devices by Older Adults. The Journal of Applied gerontology 24(5), 419–438 (2005)

9 Williams, N., Hwang, F.: Investigating Novel Point-Select techniques for Older People. In: Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics, Noordwjk, The Netherlands, pp. 614–618 (2007)

10 Czaja, S.J., Lee, C.C.: Designing computer system for older adults. In: Jacko, J., Sears, A. (eds.) Handbook of Human-Computer Interaction, pp. 413–428. Lawrence Erlbaum and Associates, New York (2003)

11 Feldmann, H.H., Woodward, M.: Evolution de la maladie d'Alzheimer. Neurology 65, S10–S17 (2005)

12 Folstein, M.F., Folstein, S., McHugh, P.R.: Mini Mental State: a practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198 (1975)

13 Hollinworth, N., Hwang, F.: Relating computer tasks to existing knowledge to improve accessibility for Older people. In: ASSETS 2010 Copright 2010, October 25-27, pp. 147–154. ACM, New York (2010)
14 Giannakouris, K.: Statistics in focus. Office for Official Publications of the European Communities. Catalogue number: KS-SF-08-072-EN-N (2008), http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-08-072/EN/KS-SF-08-072-EN.PDF
15 Lopez-Arrieta, J.M., Birks, J.: Nimodipine for primary degenerative, mixed and vascular dementia. Cochrane Database Syst. 3, CD000147 12137606 (2002)
16 Wimo, A., Winblad, B., Aguero-Torres, H., Von Strauss, E.: The magnitude of dementia occurence in the world, Alzheimer Dis. Assoc. Disord. 17(2), 63–67 (2003)

Part VI

Children and HCI

An Interface for Opportunistic Discovery of Information for Young People

Jamshid Beheshti and Andrew Large

School of Information Studies, McGill University, 3661 Peel
Montreal, Quebec H3A 1X1, Canada
{jamshid.beheshti,andrew.large}@mcgill.ca

Abstract. The exploratory study investigates a virtual reality interface for opportunistic discovery for young people. We recruited ten volunteer students, who performed a pre-assigned information search task in a VR library which consists of about 1500 web sites depicted as books organized on the shelves. We calculated a metric for opportunistic discovery of information (ODI) based on the distance of books chosen by the participants from those initially found by a conventional search. The VR interface supported ODI more than do conventional search engines.

1 Introduction

The objective of this exploratory study was to observe the information behavior of young people in a virtual reality environment designed mainly for browsing and opportunistic discovery of information. Research suggests that browsing may be a viable alternative to keyword searching for children and young adults, who might otherwise have problems in seeking information from the web (see for example, [8], [13]). Browsing, however, is not a simple act of scanning for information. Wilson devised a typology of browsing consisting of: passive attention, passive search, active search, and ongoing search [25]. Toms discussed serendipitous information retrieval in the context of browsing: "People immerse themselves in the items that interest them, meandering from topic to topic while concurrently recognizing interesting and informative information en route." They browse, scan, and read new information, the purpose of which was not their original intent [22, p.3]. De Bruijn and Spence define opportunistic browsing as: "the continuous but largely unconscious monitoring and filtering of information with the potential to trigger more purposeful behaviour [7, p. 363]. André et al [1, p. 2033] define serendipity as "the act of unexpectedly encountering something fortunate", which they regard as "valuable". They use a matrix to describe various types of serendipity, including finding information that is not relevant to the initial information need. Here, we focus on opportunistic discovery of information (ODI)[1], (or serendipitous discovery of information) which may be defined as 'information encountering' [9] that happens when one is looking for

[1] Sanda Erdelez, Guilherme DeSouza, Chi-Ren Shyu, Antonie Stam and Kevin Wise. International Workshop on Opportunistic Discovery of Information, October 21-22, 2010, Columbia, Missouri.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 541–548, 2011.

information on one topic and on the way finds information on another topic, which may or may not be relevant to the information need.

Browsing is primarily a visual activity [16], and visual-based exploratory interfaces support search activities for learning and investigating [17]. In this project, we have chosen virtual reality[2] technology as a means to facilitate visual ODI. VR environments have been used successfully in the entertainment and gaming worlds, with which our targeted audience is likely to be familiar and enjoys interacting. While young people are increasingly using electronic resources and the Internet they are still well acquainted with traditional libraries [23]. This familiarity with the library environment leads us to believe that the VR environment should represent a physical library. We chose the library metaphor to capitalize on the navigational affordances of recognized artifacts [21] over reconstructing a new model in memory [24]. The library metaphor has been used in experimental projects for more than a decade, with different degrees of success [2] [19] [20]. Zavesky et al [27, p. 619] suggest: "To fully utilize a user's inspection ability, a system must be engaging … Traditional page-based navigation is time-consuming and can be boring." For young people in particular the information environment must be engaging to ensure continual scanning and discovery of information. Engagement is also a key concept in education and in enquiry-based learning. Opportunistic discovery of information may enhance and augment the learning experience of young users of information systems.

Our research sought to answer the question: Is a virtual reality library representing the real world a more effective environment for opportunistic discovery of information for young people than a conventional information retrieval system, such as Google?

2 VRLibrary

We have developed a VR library, (henceforward called "VRLibrary") using the metaphor of a physical library with rooms, bookcases and books. The user, just as in a physical library, can walk around the library, move among the bookcases, scan the titles of books that are arranged on the bookshelves, select individual books, and open them. The difference is that the library is virtual and the books actually represent websites; when a book is "opened" it displays the contents of the web site in a window (for more details see [3]). In this environment, users can utilize *search stations* (Figure 1) located in different locations of the library to conduct conventional keyword and term searches, the results of which are displayed as red dots on a plan of the library. The user can then 'walk' through the library to spot red arrows and dots pointing to the "books" found by the search (Figure 1 and Figure 2). These features were added to the system based on the observations and results from our previous experiments [4] [5].

VRLibrary contains about 1500 links to English-language websites on Canadian history deemed to be appropriate in content and language for elementary students. The database of links was created for History Trek (www.historytrek.ca), a portal designed for children developed by an intergenerational team of researchers and children [14]. The Dewey Decimal Classification (DDC) system has been used to classify all the websites to provide a structure and organization for VRLibrary similar that used in a typical public or school library (Figure 2).

---

[2] Throughout this paper, virtual reality (VR) refers to desktop-based applications, where the environment is projected on a monitor screen and is non-immersive.

Fig. 1. VRLibrary with a *Search Station*

Fig. 2. Search results (websites) indicated by arrows

3 Conceptual Framework

Lehmann et al [15] used the degree of interest (DOI) to identify articles of potential interest in Wikipedia in a browsing context. We use a modified version of DOI to postulate a metric for the ODI itself:

$$ODI_{(v)} = \Sigma \, |V - D_{(v)}| \tag{1}$$

where V is the target website, and D is the distance of a website of interest from the target website measured in terms of information units (in this case number of websites). Although the distance $D_{(v)}$ is a function of the classification system used to organize the websites, it may be viewed as a relative measure of the browsing activity and ODI. Given the relatively small size of the database of website links in our study, we can determine the websites that are relevant for any specific information task. We can then measure the distance $D_{(v)}$ from the target website V located by the initial search to the location where opportunistic browsing or ODI has taken a user, and where other websites have been discovered. The average ODI $_{(v)}$ over all users will yield a relative measure of the browsability and opportunistic discovery of the system.

To contextualize and benchmark the metric, we calculated the highest theoretical value for ODI $_{(v)}$ for Google. Assuming that the first link to a website on the first page (screen) of the results from a Google search is a potential target website V, the maximum ODI $_{(v)}$ for the default first page containing 10 links is calculated as 45. If we include the second page of results containing 10 additional links, then the maximum ODI $_{(v)}$ is 190. However, research shows that users often do not venture beyond the first few hits, let alone the first page [11] [12]. Therefore, in practice, for Google the maximum likelihood ODI $_{(v)}$, assuming 10 links per result page is 45.

4 Methodology

4.1 Participants

Ten volunteer students, equal numbers of boys and girls ranging between 11 and 13 years old, participated in the study. A pre-test questionnaire was designed to solicit from them demographic information as well as frequencies of internet and video game usage, library visits, and information-seeking habits. All participants but one were in grades 7 and 8 (the youngest participant was in grade 6). Although they had a variety of ethno-cultural backgrounds, they were all educated in Quebec and were familiar with Canadian history. Half of the participants used the internet every day, while the other half used it at least a few time per week. They all used Google to find information for their homework and for leisure, and employed both searching and browsing strategies. The majority (7 students) used a library at least once a week, while the remainder used it at least once a month. All but two played video games, ranging in frequency from everyday to less than once per month. All were familiar with the experimental subjects chosen for the retrieval tasks.

4.2 Tasks

The experimental tasks consisted of asking participants to find information (or websites) on a topic related to Canadian history, in the context of a school project. The first task involved finding information on *holidays in Canada*. The database contains five links to websites deemed to be directly relevant to the task. For this particular topic, most of the neighbouring websites may not be directly relevant to the topic (target websites), and therefore participants should choose to ignore them. The second task was to find websites about a *parliament building*, with three relevant links to websites in the database. These tasks were chosen to represent typical topics covered in the history curriculum, and with which students were likely to be familiar. Using the DDC classification for a small collection of books resulted in dispersed clusters of subjects, which may not have been inductive to browsing. Nevertheless, it is a classification system used most often in school and public libraries.

4.3 The Procedure

Each student undertook one task independently. At the outset the purpose of the research was briefly explained to the student by the research assistant, following a script to ensure consistency across students/assistants. It was emphasized that the interfaces and not the student was being evaluated. Each student was introduced to the project with a brief demonstration of VRLibrary, showing the search and browsing features, which took about five minutes. The student was then given one of the two tasks and was asked to retrieve relevant websites in the VRLibrary. They were also asked to limit their searching and browsing to the first page of the website and not to follow the internal hyperlinks on the site. Participants' information seeking process was observed unobtrusively by a research assistant. Once students were satisfied with the information (websites) they had selected, a post-test interview was conducted to determine their impressions of the VRLibrary, and the reasons for their selections of websites other than those found in the initial search.

Four students were asked individually to locate information on the first task, *holidays in Canada*, while six were given the second task, *parliament building*. The research assistants noted the student's affective behaviour as well as any obvious problems, comments or questions raised by the student. The on-screen activity of the students as well as any spoken commentary by either student or research assistants was recorded.

4 Results

All students began with keyword or term searching on one of the search stations in VRLibrary. In practice it is difficult, if not impossible, to begin browsing immediately before identifying a starting point in the collection. On Task 1, finding information on *holidays*, three of the four students managed to locate (by clicking on the book spines) the five target websites with relative ease. The first student continued browsing the shelves after finding the target websites and clicked on three additional books (and opened the websites) before going back to the search station. The distances between these additional books (websites) and the closest target websites were calculated based on the order of the books arranged in DDC in the database. We assumed that any book that was chosen directly as a result of a keyword/term search on the search stations should not be counted as an ODI act, and therefore did not include them in our calculations. The next two participants, while browsing the shelves, did not click on any other books but the target websites, and were satisfied with their search results. The last participant to perform Task 1, also browsed and clicked on several books that were relatively far from the target websites. The ODI for Task 1 was calculated using the formula (1), and over the four participants averaged 156 books (links to websites).

Six students performed Task 2, finding information on *parliament building*. The ODI activity for each participant varied significantly, ranging from 0 to 589 books. The average ODI $_{(v)}$ for Task 2 was 168 books.

Participants' subjective evaluation of the system, captured through the post-test interviews, showed a very high level of satisfaction with VRLibrary. The young people in the study used terms such as 'cool', and 'easy to use' to describe VRLibrary, and found the environment very appealing and engaging.

5 Discussion and Conclusions

Two experimental tasks were performed by 10 participants to explore the opportunistic discovery of information (ODI) of a virtual reality library. A new metric was conceived to quantify ODI, defined as the sum of distances of the links to websites of interest from the target websites retrieved in a conventional search. The ODI for one task was calculated as 156 links and for another task was 168 links. Equivalent tasks on Google show a maximum theoretical ODI of 45 links, assuming young users do not browse past the first result page. These results suggest that VRLibrary with average ODIs higher than the theoretical value calculated for Google may inspire young people to discover information more readily than conventional search engines.

In one instance during the experiment, a student chose through ODI two completely unrelated books because she was "curious". The element of curiosity or surprise is a prerequisite for ODI [18]. In one study, Bilal found that for self-generated tasks, most children in her experiment visited websites that were irrelevant to their task, and concluded that they likely wanted to "explore topics of interest other than those they were pursuing" [6, p. 1177]. While we did not build any triggering mechanism for discovery in VRLibrary, as recommended by some researchers [1] [10], it seems that the system supports and perhaps encourages ODI.

Caution must be exercised in generalizing from one small-scale study using one interface. First, the observed wide variations among the participants in searching and browsing in the VRLibrary may suggest different cognitive styles in information behaviour among the young users. Westerman et al [25] found in their study of testing a two and three-dimensional virtual information space that the adult participants with lower cognitive ability out-performed others in the three-dimensional environment, in terms of recall and precision, and performance efficiency. To use quantifiable measures and a metric such as the ODI $_{(v)}$, larger samples may produce more precise results in experimental studies. Second, the organization of the information (or the links), may affect opportunistic discovery. In the VRLibrary, we used the Dewey

Decimal Classification to organize a small collection of links on Canadian history to mimic the organizational structures of many real-world public and school libraries. Other clustering methods, such as automatic text categorization (TC) using machine learning algorithms, or self-organizing maps [20] may have different effects on ODI. Although logistically not feasible, for comparative analysis a similar algorithm as in Google's Page ranking should be utilized to organize the links in the VRLibrary to measure ODI. Finally, McCay-Peet and Toms discuss the challenges and tribulations of using experimental laboratory goal-oriented methods to measure ODI [18]. Ideally, a field study under operational setting, where students would have the option of using VRLibrary for completion of classroom projects, will be the ultimate test for measuring ODI in a virtual environment.

Despite the limitations of the study, however, a common theme emerged from the experiment. VRLibrary's basic premise of using a real-world metaphor seems to be an effective and exceptionally engaging method of presenting information to young users for opportunistic discovery of information.

Acknowledgements. The VRLibrary program was developed by Ian Clement; his expertise is gratefully acknowledged, as is the help of our research assistants, Nouf Khoshman and Megan Stecyk. The research could not have been conducted without the student volunteers. This project was funded by the Social Sciences and Humanities Research Council (Canada).

References

1. André, P., Teevan, J., Dumais, S.T.: From X-rays to silly putty via Uranus: Serendipity and its role in web search. In: CHI 2009: Proceedings of the 27*th* International Conference on Human Factors in Computing Systems, April 9. ACM, New York (2009)
2. Beheshti, J., Large, V., Bialek, M.: Public Access Catalogue Extension (PACE): A Browsable Graphical Interface. Information Technology and Libraries 15(4), 231–240 (1996)
3. Beheshti, J., Large, A., Julien, C.A.: Designing A Virtual Reality Interface for Children's Web Portals Data, Information, and Knowledge in a Networked World. In: Annual Conference of the Canadian Association for Information Science, London, Ontario, June 2-4 (2005)
4. Beheshti, J., Large, A.: Preliminary Design Indicators to Desktop Virtual Reality Environments. In: ASIST 2006 Annual Meeting, Information Realities: Shaping the Digital Future for All. SIG HCI Research Symposium: Human-Computer Interaction in Information Intensive Environments, Austin, Texas, November 5 (2006)
5. Beheshti, J., Large, A., Clement, I., Tabatabaei, N.: Evaluating the usability of a virtual reality information system for children Information Sharing in a Fragmented World: Crossing Boundaries. In: Annual Conference of the Canadian Association for Information Science, Montreal, Quebec, May 10-12 (2007)
6. Bilal, D.: Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated search tasks. Journal of the American Society for Information Science and Technology 53(13), 1170–1183 (2002)
7. de Bruijn, O., Spence, R.: Serendipity within a ubiquitous computing environment: A case for opportunistic browsing. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 362–369. Springer, Heidelberg (2001)

8. Druin, A., Bederson, B.B., Hourcade, J.P., Sherman, L., Revelle, G., Platner, M., Weng, S.: Designing a digital library for young children: an intergenerational partership. In: Proceedings of JCDL, pp. 398–405. ACM Press, New York (2001)
9. Erdelez, S.: Information encountering: It's more than just bumping into information. Bulletin of the American Society for Information Science, 25–29 (Febrary/March 1999)
10. Erdelez, S.: Investigation of information encountering in the controlled research environment. Information Processing and Management 40(6), 1013–1025 (2004)
11. Guan, Z., Cutrell, E.: An eye tracking study of the effect of target rank on web search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York (2007)
12. Joachims, T., Granka, L.: Accurately Interpreting Clickthrough Data as Implicit Feedback. In: ACM SIGIR 2005, pp. 154–161 (2005)
13. Large, A., Beheshti, J.: The Web as a classroom resource: reactions from the users. Journal of American Society for Informational Science 51(12), 1069–1080 (2000)
14. Large, A., Beheshti, J., Nesset, V., Bowler, L.: Children's Web Portals: Can an Intergenerational Design Team Deliver the Goods? In: Chelton, M.K., Cool, C. (eds.) Youth Information-Seeking Behavior, vol. 2, pp. 279–311. Scarecrow, Lantham (2007)
15. Lehman, S., Schwanecke, U., Dörner, R.: Interactive visualization for opportunistic exploration of large document collections. Information Systems 35(2), 260–269 (2010)
16. Marchionini, G.: Information seeking in electronic environments. The Press Syndicate of the University of Cambridge, Cambridge (1995)
17. Marchionini, G.: Exploratory search: From finding to understanding. Communications of the ACM 49(4), 41–46 (2006)
18. McCay-Peet, L., Toms, E.G.: The Process of serendipity in knowledge work, August 18-21. ACM, New Brunswick (2010)
19. Pejtersen, A.M.: A Library system for information retrieval based on a cognitive task analysis and supported by an icon-based interface. In: SIGIR 1989 Proceedings of the 12*th* Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 40–47 (1989)
20. Rauber, A., Merkl, D.: The SOMLib digital library system. In: Abiteboul, S., Vercoustre, A.-M. (eds.) ECDL 1999. LNCS, vol. 1696, pp. 323–342. Springer, Heidelberg (1999)
21. Shiaw, H., Jacob, R.J.K., Crane, G.R.: The 3D vase museum: a new approach to context in a digital library. In: Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, June 7-11, pp. 125–134 (2004)
22. Toms, E.: Serendipitous information retrieval. In: Proc. of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries (2000)
23. Walter, V.A.: Public library service to children and teens: A research agenda. Library Trends 51(4), 571–589 (2003)
24. Ware, C.: Visual thinking for design. Morgan Kaufmann, Burlington (2008)
25. Westerman, S.J., Collins, J., Cribbin, T.: Browsing a document collection represented in two- and three-dimensional virtual information space. International Journal of Human-Computer Studies 62(6), 713–736 (2005)
26. Wilson, T.D.: Information behaviour: an interdisciplinary perspective. Information Processing and Management 33(4), 551–572 (1997)
27. Zavesky, E., Chang, S.F., Yang, C.C.: Visual Islands: Intuitive browsing of visual search results. In: CIVR 2008: Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, Niagara Falls, Ont, July 7-9. ACM Press, New York (2008)

Evaluating Leading Web Search Engines on Children's Queries

Dania Bilal and Rebekah Ellis

School of Information Sciences
University of Tennessee, 1345 Communications 451, Knoxville, TN 37996
{dania,rellis17}@utk.edu

Abstract. This study compared retrieved results, relevance ranking, and overlap across Google, Yahoo!, Bing, Yahoo Kids!, and Ask Kids on 15 queries constructed by middle school children. Queries included one word, two words, and multiple words/phrases/natural language, and the results were benchmarked against Google and Yahoo Kids! top 5 and top 10 retrieved results using a new relevance ranking metric. Yahoo! and Bing yielded similar results on all queries, but their relevance ranking differed on one-word queries. Ask Kids outperformed Yahoo Kids! on all queries, and a modest percentage of results had the same relevance ranking as Google. Yahoo Kids! and Ask Kids returned unique results that were not retrieved by the other three engines on the first results page. Yahoo! and Bing produced the highest percentage in overlap with Google followed by Ask Kids. Implications are made for children and mediators concerning the use of search engines on children's queries.

Keywords: Children, queries, query construction, web search engines, evaluation, information retrieval, relevance ranking, ranking comparison, overlap, unique results, Google, Yahoo!, Bing, Yahoo Kids!, Ask Kids.

1 Introduction

Children use the Web on a daily basis to support their information needs, find materials for research projects, and communicate with others using social media. A recent study shows that the time children aged 2-11 spend online increased 63% between 2004 and 2009 [1]. Children's use of Web search engines accounts for a portion of this time increase. Today's children rank Google as their top choice for finding information, followed by Yahoo!, Bing, and Ask.com [2]. Previous studies of Google, Yahoo!, Live Search/MSN, and AskJeeves revealed inconsistencies in retrieved results, overlap [3][4], and a high percentage (84.9%) of results that were unique to one specific engine [5]. To a child, this means that certain relevant results to a query will most likely be missed if the child uses only one search engine. As children are increasingly using leading search engines designed for the general public (Google, Yahoo!, Bing) rather than engines that are specifically designed for their age levels (i.e., Yahoo Kids! and Ask Kids), we need to develop an understanding of how these search engines differ in retrieval and relevance ranking of results for children's

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 549–558, 2011.

queries. Such an understanding is significant because these engines are not equal; they have their own search capabilities, ranking algorithms, limitations, and complexity. As Thelwall notes, "although search engines are entirely logical because they are computer programs, their complexity means that the results they present often have inconsistencies and unexpected variability" [3]. Results gained from this study should be useful for children and their mediators (parents, teachers, practitioners) in determining which search engines to use for child-driven content on specific types of queries (one word, two words, and multiple word/phrases/natural language). In addition, the results could help generate improvements in the interface design of these search engines in support of children's effective Web interaction.

2 Related Studies

2.1 Evaluation of Search Engines Designed for the General Public

In a study of 1,587 single-word queries, Thelwall [3] compared Google, Yahoo!, and Live Search (previously MSN) for hit counts, consistencies, and matching URLs. Google and Yahoo! provided five to six times more hits than Live Search, and Yahoo! returned slightly more matching URLs than Google. The engines were more consistent in the top-level domains represented in the URLs, although Yahoo! returned the most. For hit counts, Google outperformed Yahoo! and Live Search, but for the range of domains and sites represented in the results, Yahoo! surpassed the other two engines. Lewandowski [4] compared the retrieval performance of five Web search engines (i.e., Yahoo!, Ask, Google, MSN, and Seekport) on single-word queries. He found that Google and Yahoo! outperformed the others, but neither was found to be superior overall to the other, although Google held a slight advantage when only the top three results were considered. In an earlier study, Spink, Jansen, Blakely, and Koshman [5] compared retrieved results from Google, Yahoo!, MSN, and Ask Jeeves. They found that 84.9% of the retrieved results from the first results page were unique to 1 of the 4 search engines and that the percentage of unique results was higher than the percentage of results that overlapped across the engines.

2.2 Evaluation of Search Engines Designed for Children

The earliest study that evaluated the retrieval performance of search engines designed for children was conducted by Bilal [6]. She compared Yahooligans!, Ask Jeeves Kids, and Super Snooper (now a defunct engine) on a set of children's queries (single words, multiple words, and natural language) using the following four criteria: retrieval output, relevance, overlap, and redundancy. Yahooligans! was effective on queries with single terms and ineffective on queries with multiple terms or natural language. Ask Jeeves Kids retrieved results from Yahooligans! for queries using natural language that the latter failed to retrieve. There was an overlap in returned results between Yahooligans! and Ask Jeeves Kids.

Taking a different approach, Large, Beheshti, and Rahman [7] evaluated Yahooligans!, Ask Jeeves Kids, KidsClick, and Lycos Zone from the perspective of middle school children. The authors generated four child-centered design criteria:

goals, visual design, information architecture, and personalization that they considered essential for implementation in the design of children's search engines. In addition, they found that the four engines were more effective on single-word searches than on queries with phrases or natural language.

Bilal [8][9][10] evaluated Yahooligans! based on middle school children's information-seeking behavior on fact-based, research-based, and fully self-generated tasks. Overall, children were more successful on the fully self-generated task than on the other two tasks. Additionally, children clicked more on results displayed near the top of the first results page than in the middle or bottom of the page. Children's use of single-word queries was more effective than queries with multiple words or natural language. Most of the breakdowns children experienced resulted from the inadequate interface design of Yahooligans!

In a recent study, Druin, et al. [11] explored the keyword searching of children ages 7, 9, and 11 in using Google on four assigned tasks. A small group of children demonstrated strong search expertise and very few children were successful in formulating complex queries. Seven search roles were identified based on children's information seeking. Children's success varied depending on search role. For example, p*ower searchers* were the most successful in accessing information and in assessing whether they had found what they needed.

In summary, previous studies of search engines, namely Google, Yahoo!, MSN/Live Search (now Bing), and Ask Jeeves (now Ask) uncovered inconsistencies in the range of URLs, domain coverage, relevance ranking, and overlap across the engines. Studies of children's information behavior on the Web revealed that they were more successful in finding information from simple queries than for complex queries. One study, evaluated the retrieval performance of three child-driven search engines. However, this study is dated in the late 1990s. Today, children prefer to use Google and like engines over Yahoo Kids! and Ask Kids. Nevertheless, there has been little comparison between these two search engine groups on children's queries. This study begins to examine differences and similarities between these engine groups in retrieving and ranking results on the first results page for queries with one word, two words, and multiple words/phrases/or natural language.

3 Research Questions

This study addressed these research questions:

1. Using Yahoo Kids! as the benchmark, what are the differences in relevance ranking of the top five retrieved results on the first results page across Google, Yahoo!, Bing, and Ask Kids on children's queries (one word, two words, and multiple words/phrases/natural language)?
2. Using Google as the benchmark, what are the differences in relevance ranking of top five retrieved results on the first results page across Yahoo!, Bing, Yahoo Kids!, and Ask Kids on children's queries (one word, two words, and multiple words/phrases/natural language)?
3. Using Google as the benchmark, what is the percentage of overlap in retrieved results on the first results page for the top 5 and top 10 results across Yahoo!, Bing,

Yahoo Kids!, and Ask Kids on children's queries (one word, two words, and multiple words/phrases/natural language)?

4 Method

This study employed the quantitative method for comparing children's queries across the five search engines (Google, Yahoo, Bing, Yahoo Kids!, and Ask Kids).

Query Set. We examined the published literature from 1989 to present on children's interaction with digital tools (e.g., OPACs, CD-ROMs, and search engines). We identified 130 tasks that were assigned to children and/or self-selected by them, and we focused on studies of children in grade levels 5-9. For these tasks, we examined how children queried a given digital tool (i.e., search statements using one word, two words, multiple words/phrases/natural language) so that we could query each of the five search engines using these statements.

Query Sample. We selected five queries with single words, five with two words, and five with multiple words/phrases/natural language (Table 1) to use in this study.

Table 1. Queries submitted to search engines

One-word queries	Two-word queries	Multiple words/phrases/ natural language queries
Ozone Diabetes The Simpsons Spiders Vegetarians	Spanish armada Speed skating Endangered animals Ancient numerals Olympic hockey	Why dolphins migrate? Women in space Clock using sun and stick What are the three most common crimes in California? Women at war

Data. We submitted each query to each search engine, retrieved the results on the first results page, printed out the results, and coded and analyzed them manually. Each set of queries was submitted within a minute of each other, and all 15 queries occurred on January 31, 2011, between 10:00 and 11:30 p.m. to avoid any occurrence of possible changes in retrieved results due to search engine updates.

We used Yahoo Kids! and Google as benchmarks for comparing the retrieved results, relevance ranking of the results, and overlap across the engines. Yahoo Kids! was selected as a benchmark because it is specifically designed for children ages 7-12, and Google was selected as a benchmark because it is the most used by children.

The first results page on a given query in a given search engine was printed out and the top five retrieved results were considered for comparison against the benchmark. We coded the top 5 ranked results retrieved by Yahoo Kids! and Google using the codes, 1-5 (1=first ranked result; 5=fifth ranked result). Our submission of the 15 queries to the 5 search engines resulted in a total of 75 printouts of retrieved results (25 printouts for queries with 1 word, 25 for queries with 2 words, and 25 for queries with multiple words/phrases/natural language). We excluded advertisements from the results each search engine retrieved. We employed a *relevance ranking metric* with four values: 0=same ranking as the rank of the benchmark; (+n) (e.g., +1,+2,+3) means that the retrieved result by a given search engine is ranked 1, 2, or 3 positions (as applicable) below the rank of a retrieved result by the benchmark; (-n) (e.g., -1, -2,-3) means that the retrieved result is ranked 1, 2, or 3 positions (as applicable) above the retrieved result by the benchmark; and NR means that no results were retrieved for a given query by a given search engine on the first results page that matched with the top 5 or top 10 results retrieved by a benchmark. We coded the 75 printouts of retrieved results manually using the values explained above and shown in Table 2. We compared the ranking position of the results retrieved for a given query by a given search engine against the top five ranked results retrieved by the benchmarks, one time using Google as the benchmark and another time using Yahoo Kids! as the benchmark.

Table 2. Values for ranking of retrieved results across search engines

0, where a result has the same ranking position as the ranking by the benchmark.
+n, where n=ranking position below the benchmark rank. For example, a result with a +1 means it was ranked one position below the ranking by the benchmark.
-n, where n=ranking position above the benchmark rank. For example, a result with a -1 means it was ranked one position above the ranking by the benchmark.
NR = Not retrieved.

5 Results

In reporting the results, we use the code (YK-R) to indicate Yahoo Kids! relevance ranking and the code (G-R) to denote Google relevance ranking of a given retrieved result.

Differences in Relevance Ranking of Retrieved Results across Search Engines – Yahoo Kids as Benchmark on One-Word Queries. No results were retrieved by Google, Yahoo!, and Bing that matched with the top five results retrieved on one-word queries by Yahoo Kids! Ask Kids retrieved one matching result for the query *diabetes*, but it did not have the same ranking as Yahoo Kids (+3 vs. YK-R=3, respectively).

Differences in Relevance Ranking of Retrieved Results across Search Engines – Google as Benchmark on One-Word Queries. On the query *ozone*, Yahoo Kids! did not return any results on its first results page that matched with the top five results retrieved by Google. Two results retrieved by Yahoo! and two retrieved by Bing were identical and had the same ranking as Google (0 and G-R=1, respectively) and

(0 and G-R=3, respectively). Ask Kids retrieved two matched results, but their ranking was different from Google (-2, +5 vs. G-R=3 and G-R=4, respectively). For the query *diabetes*, Yahoo! retrieved three results that matched with Google, but only two had the same ranking as Google (0, 0 and G-R=2, G-R=3, respectively). Ask Kids retrieved only one result that matched with Google, but it did not have the same ranking as Google (-1 vs. G-R=2, respectively). For the query *spiders*, Yahoo! and Bing retrieved three results each that matched the results Google retrieved, but only one result had the same ranking as Google (0 and G-R=1, respectively). For the query *vegetarians*, Yahoo! and Bing retrieved four results each that matched with Google, but only one result that Yahoo! retrieved had the same ranking as Google (0 and G-R=1, respectively). Ask Kids retrieved one matched result, but it did not have the same ranking as Google (-3 vs. G-R=4, respectively). For the query *the Simpsons*, Yahoo! retrieved four results that matched with and had the same ranking as Google. Bing retrieved four results of which only two had the same ranking as Google.

Differences in Relevance Ranking of Retrieved Results across Search Engines – Yahoo Kids! as Benchmark on Two-Word Queries. On the query *Spanish armada*, Google retrieved four results that matched with Yahoo Kids!, but they did not have the same ranking as Yahoo Kids! (+6, +2, +5,+1 vs. YK-R=1, YK-R=2, YK-R=3, YK-R=4, respectively). Yahoo! and Bing retrieved one identical result on this query that matched with Yahoo Kids!, but it did not have the same ranking as Yahoo Kids! (+6 vs. YK-R=2). Ask Kids also retrieved one matched result that did not have the same ranking as Yahoo Kids! (+1 vs. YK-R=1, respectively). For the query *speed skating*, only Ask Kids retrieved two matched results but they did not have the same ranking as Yahoo Kids! (+3, +2 vs. YK-R=1, YK-R=2, respectively). None of the four engines returned results that matched with Yahoo Kids! for the query *ancient numerals* or *Olympic hockey*.

Differences in Relevance Ranking of Retrieved Results across Search Engines – Google as Benchmark on Two-Word Queries. Yahoo! and Bing retrieved four identical results each for the query Spanish armada, that matched with Google of which one result had the same relevance ranking as Google (0 and G-R=1). Yahoo Kids! retrieved two results that matched with Google, but none of them had the same ranking as Google.

For the query *endangered animals*, Yahoo! and Bing retrieved one result each that had the same ranking as Google (0 and G-R=2). Ask Kids retrieved three matched results, but none of them had the same ranking as Google. For the query *speed skating*, Yahoo! and Bing retrieved three matched results, but none of them had the same ranking as Google. Yahoo Kids! did not retrieve any results that matched with Google. Ask Kids retrieved one result that matched with and had the same ranking as Google (0 vs. G-R=1).

For the query *ancient numerals*, Yahoo! and Bing retrieved three identical results each that matched with Google, but none of them had the same ranking as Google (-2, +6, +4 vs. G-R=3, G-R=4, G-R=5, respectively). Ask Kids retrieved only one matched result, but it did not have the same ranking as Google (+6 vs. G-R=5). Yahoo Kids! retrieved no results from its first results page that matched with Google.

For the query *Olympic hockey*, Yahoo! and Bing retrieved three identical results each of which one result had the same ranking as Google (0 vs. G-R=1). Neither Yahoo Kids! nor Ask Kids retrieved results that matched with Google.

Differences in Relevance Ranking of Retrieved Results across Search Engines – Yahoo Kids! as Benchmark on Multiple Words/Phrases/Natural Language Queries. No engine retrieved results on its first results page that matched with Yahoo Kids!, with the exception of Google that returned one result for *the query women in space*, but it did not have the same ranking as Yahoo Kids! (+7 vs. YK-R=1, respectively).

Differences in Relevance Ranking of Retrieved Results across Search Engines – Google as Benchmark on Multiple Words/Phrases/Natural Language Queries. For *the query women in space*, Yahoo! and Bing retrieved four identical results each that matched with Google, but they did not have the same ranking as Google. Ask Kids retrieved three results that matched with Google, but only one result had the same ranking as Google (0 and G-R=4; +1, -2 vs. G-R=5, G-R=1, respectively). Yahoo Kids! did not retrieve any result that matched with Google.

For another query, *women at war*, Yahoo! and Bing retrieved one identical result that had the same ranking as Google (0 and G-R=1) and one result that matched with but did not have the same ranking as Google (+1 vs. G-R=3). Bing retrieved one result that had the same ranking as Google (0 and G-R=3). Yahoo Kids! retrieved one result that matched with Google, but it did not have the same ranking as Google (-2 vs. G-R=4) and Ask Kids retrieved one matched result that had the same ranking as Google (0 and G-R=3). For the query *clock using sun and stick*, Yahoo! and Bing retrieved one identical result but it did not have the same ranking as Google (-1 vs. G-R=3).

For the query *what are the three most common crimes in California*, no engine retrieved results that matched with Google. As to the query *why do dolphins migrate*, Yahoo! and Bing retrieved one identical result that had the same ranking as Google (0 vs. G-R=3), and three additional results that did not have the same ranking as Google (+1, +4, -3 vs. G-R=1, G-R=2, G-R=4, respectively). Both Yahoo Kids! and Ask Kids did not retrieve results on the first results page that matched with Google on these two queries.

5.1 Overlap in Results Across the Search Engines

We assessed the overlap in the top 5 and top 10 results retrieved by each search engine against Google's ranking of the top 5 retrieved results for a given query. To calculate the overlap in the top 10 results retrieved, we counted the results that matched with Google on each query by each search engine and totaled the matched results across all 5 queries by query type (one word, two words, and multiple words/phrases/natural language). We divided the total number of matched results for each query type by 50, where 50 is the maximum number of first results pages a given search engine would retrieve on the 5 queries by query type (10 results per page for a given query multiplied by 5 result pages).

Overlap in the top five and top ten ranked results.
One-word queries: Both Yahoo! and Bing had the same percentage in overlap with Google (36%) in the top 5 ranked results, followed by Ask Kids (12%). There was no overlap between Yahoo Kids! and Google on these queries. Yahoo! produced the highest percentage in overlap (84%) in the top 10 retrieved results benchmarked against Google, followed by Bing (36%) and Ask Kids (12%). Yahoo Kids! did not retrieve results that matched with Google, yielding no overlap (Table 3).
Two-word queries: Both Yahoo! and Bing resulted in an equal percentage in overlap with Google, 34% in the top 5 and 38% in the top 10 results, followed by Ask Kids, 22% in the top 5 and 32% in the top 10. Yahoo Kids! produced the lowest percentage in overlap with Google in both the top 5 and the top 10 (0.4%) (Table 3).
Multiple words/phrases/natural language queries: Both Yahoo! and Bing yielded the same percentage in overlap with Google, 22% in the top 5 and 40% in the top 10 results. Ask Kids resulted in 10% overlap with Google in the top 5 and 16% in the top 10. Yahoo Kids! showed (0.2%) overlap with Google in both the top 5 and top 10 results. Here too, Ask Kids surpassed Yahoo Kids! in finding matches that overlapped with Google.

Table 3. Overlap with Google across the search engines on children's queries

Engine	One-word queries Percent* in overlap		Two-word queries Percent* in Overlap		Multiple words/ phrases/natural language queries Percent* in overlap	
	Top 5	Top 10	Top 5	Top 10	Top 5	Top 10
Yahoo!	36%	84%	34%	38%	22%	40%
Bing	36%	36%	34%	38%	22%	40%
Yahoo Kids!	-0-	-0-	0.4%	0.4%	0.2%	0.2%
Ask Kids	12%	12%	22%	32%	10%	16%

*Percentage of overlap is based on the first results page retrieved by each engine on the five queries submitted.

6 Discussion and Conclusion

One of the key findings of this study is that using Google as a benchmark against which to compare the retrieved results and their relevance ranking across the four engines was more effective than using Yahoo Kids! as a benchmark. In fact, Yahoo Kids! is *unique* in its content, especially since it is targeted to children ages 7-12, as described in its online help. Conversely, Ask Kids which is also designed for children surpassed Yahoo Kids! in retrieving a number of results that not only matched with Google, but also had the same relevance ranking as Google. The difference in retrieval between Yahoo Kids! and Ask Kids may be attributed to the level of indexing the search engine employed and to the fact that unlike Yahoo!, Yahoo Kids!

is more of a directory than a search engine, and, therefore, its indexing methodology is less comprehensive than that employed by Ask Kids.

The overlap with Google across the four search engines varied by query type. Yahoo! and Bing retrieved identical results on most queries and had the highest percentage of overlap with Google. However, the relevance ranking of results between these two engines was different, especially on one-word queries. It appears that these engines have a common indexing methodology but use different ranking techniques.

Another key finding is that the percentage of results that had the same relevance ranking as Google across the four engines was the highest on one-word queries and the lowest on multiple word/phrases/natural language queries. This finding is congruent with the results of previous studies on search engines designed for the general public [3][4][5] and those designed for children [8][9][10], which revealed that the engines were more effective on single-word queries than on two-words or natural language queries. This is an area of research that needs further investigation to uncover these underlying retrieval problems across these engines.

Although it is designed for children, Ask Kids! outperformed Yahoo Kids! in retrieving results that overlapped with Google on all types of queries. This finding may be attributed to the customized type of documents indexed by Yahoo Kids! that tend to be highly filtered and judged based on the reading abilities of elementary and middle school children. This finding could mean that children and their mediators may want to use Ask Kids over Yahoo Kids! to find child-centric information for their queries without sifting through the many results retrieved by Google.

Another finding is that both Yahoo Kids! and Ask Kids retrieved *unique* results that were not shared by Google, Yahoo!, or Bing on the first results page for the queries. These results are child driven, and, therefore, children and their mediators may miss exposure to these results that could be relevant to the queries by focusing only on Google or like engines. Due to the fact that we focused on system-driven relevance judgment and ranking of retrieved results across the five engines, we did not judge relevance of the *unique* results that both Yahoo Kids! and Ask Kids retrieved on the queries. In addition, we did not judge relevance of the results retrieved by Google, Yahoo!, and Bing against children's reading abilities or reading comprehension.

This study provided an understanding of differences and similarities between search engines designed for adult users and engines designed for children. It uncovered strengths and weaknesses of retrieved results, relevance ranking, and overlap across the engines on different types of children's queries. Much remains to be learned about these search engines to optimize their use by children. Some questions to address in future studies are: 1. What are the search capabilities and retrieval performance of Yahoo Kids! and Ask Kids on different types of children's queries? 2. Of the five search engines used in this study, what engine is the "best" on children's queries in specific subject domains (e.g., science, social science), and 3. To what extent does the system-driven relevance ranking of retrieved results across the five engines vary from the user-based relevance ranking of the same results on different types of queries?

References

1. Nielsen, J.: Children's Websites: Usability issues in designing for kids, http://www.useit.com/alertbox/childen.html
2. Olsen, S.: Google seeks to help children search better, http://www.nytimes.com/2009/12/26/technology/internet/26kidsearch.html
3. Thelwall, M.: Quantitative comparisons of search engine results. J. Am. Soc. Inform. Sci. 59, 1702–1710 (2008)
4. Lewandowski, D.: The retrieval effectiveness of Web search engines: Considering results descriptions. J. Doc. 64, 915–937 (2008)
5. Spink, A., Jansen, B.J., Blakely, C., Koshman, S.: A study of overlap and uniqueness among major Web search engines. Inform. Process. Manag. 42, 1379–1391 (2006)
6. Bilal, D.: Web search engines for children: A comparative study and performance evaluation of Yahooligans!, Ask Jeeves Kids, and Super Snooper. In: Kids, J., Snooper, S. (eds.) 62nd Am. Soc. Inform. Sci. Meeting, pp. 70–82. Information Today, Inc., Medford (1999)
7. Large, A., Beheshti, J., Rahman, T.: Design criteria for children's Web portals: The users speak out. J. Am. Soc. Inform. Sci. Tech. 53, 79–94 (2002)
8. Bilal, D.: Children's use of the Yahooligans! Web search engine. I. Cognitive, physical, and affective behaviors on fact-based tasks. J. Am. Soc. Inform. Sci. 51, 646–665 (2000)
9. Bilal, D.: Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated tasks. J. Am. Soc. Inform. Sci. Tech. 53, 1170–1183 (2002)
10. Bilal, D.: Children's use of the Yahooligans! Web search engine. II. Cognitive and physical behaviors on research tasks. J. Am. Soc. for Inform. Sci. Tech. 52, 118–137 (2001)
11. Druin, A., et al.: How children search the internet with keyword interfaces. In: 8th International Conference on Interaction Design and Children, Como, Italy, pp. 89–96 (2009)

How Children Can Design the Future

Mona Leigh Guha[1], Allison Druin[1], and Jerry Alan Fails[2]

[1] Human Computer Interaction Lab, University of Maryland, USA
[2] Montclair State Univerity, USA
mona@cs.umd.edu, allisond@umiacs.umd.edu,
failsj@mail.montclair.edu

Abstract. Over the past 15 years, children have become more integrally involved in the design of their technology. In this paper, we present the idea that design partnering methods, specifically Cooperative Inquiry, used for designing technology with children can and should now be extended into informal and formal educational settings.

Keywords: Children, Cooperative Inquiry, Design Process, Education.

1 Introduction

Over the past fifteen years, there has been a proliferation worldwide of research on technology design processes with children. Although historically children have been involved in design processes as testers and users, today it is becoming more common for children to be involved in more long-term and participatory roles as informants and design partners [1]. Our recent research [2], showed design partners have positive social and cognitive experiences during the design process. We believe it is time to extend the role of design partners from technology design processes into other informal and formal educational settings.

In choosing to work with children in the design process, designers of children's technology need to consider which design method is a good match for their team. In making this decision, designers must consider the unique needs of their team, as well as their goals. Considerations may include the amount of time adults designers can work with children, the cost of supplies for working in this manner, and the research questions being asked. If a designer considers the social and cognitive experiences of children to be important then choosing to work with children as design partners may not only benefit the technology created, but also has the potential to provide positive experiences to the children involved in the design process [2].

We hope that the number of designers choosing to work with children as design partners in the design of new children's technology will increase. However, we believe that it is time for child design partners to go beyond designing just technology. We believe deeply in the power of children designing their own future. We have repeatedly demonstrated this through our commitment to designing technology with and for children. We believe that these experiences can be extended and enhanced through long-term involvement of children as design partners in creating their own educational experiences.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 559–569, 2011.

Having worked for years with child design partners [1-5], and having seen the potential for positive cognitive and social experiences for children participating with these methods, we believe that now is the time to not only allow children to design the technology of their future, but to let them design the education of their present. This paper will present an application of Human-Computer Interaction research done over the past fifteen years into practical innovation of education reform.

2 Designing Technology for Children

2.1 Children's Roles in the Technology Design Process

The past years have seen a proliferation of designers working with children in the design of new technology for children. There are many ways in which children can participate in the design of new technology. Our work follows the taxonomy set forth by Druin [1] as illustrated in **Fig. 1**. In this conceptualization, there are four main roles that children can play in the design of their technology: user, tester, informant, and design partner. The growing rings imply that as one moves from user to design partner, each role encompasses the one before it and becomes wider and more involved.

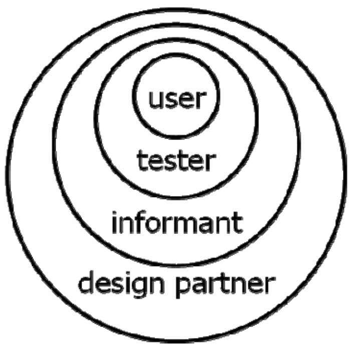


Fig. 1. Roles children can play in the technology design process

The most historic role for children to play in the design process is as users. In this role, children are given technology to use once it has been fully designed, developed, and generally deployed. While this role has benefits, including low overhead in terms of financial and human resource support, the main drawback to this role is that children also have very little input into the process. As users, children may be able to play with the technology and tell or show designers what they like and do not like about it, however, there is generally very little room for revision of the technology at the point that the children use it.

A slightly more involved role is as tester, in which children are brought in to test out technology earlier in the design process. The advantage to this over working with children as users is that designers can gather input from the children earlier in the design and development cycle when significant revisions to the technology can still be made. This role is appealing to many designers as it will afford input earlier in the process and thus real changes can be made to technology before it is deployed.

There is a qualitative shift between testers and the next role, that of informants [6]. The roles of users and testers are, to an extent, passive. Children are given technologies that adults have designed and developed and are asked to play with them. While this input is certainly useful, it is not until the level of informant where children are seen as potential problem-solvers in the technology design process. As informants, children can be called in at any time during the design process. Possibly a technology design team has come to an impasse during the brainstorming phase and decides to ask children to come help with this process. Or, further along in the process, the team is unsure how the interface of a piece of software should function and at that point asks for the informants to come in to help. Thus, child informants are thought of as more than end-users, rather, it is at this level that designers begin to think of children as able to contribute to design.

Moving one step further to the most involved level, children who are design partners are considered to be equal participants in the design process with the adult designers. These children attend design sessions regularly and participate in all aspects of design, from initial brainstorming to interface design to aiding in testing with users. This method has many advantages, including gaining input from children at earlier phases in the design cycle which may lead to fewer revisions before reaching a final product [5, 7]. Children's voices are heard and respected throughout the design process.

Other than user, tester, informant, and design partner, there are other methods for including children in the technology design process. In Bonded Design [8, 9], children participate for a short-term but intensive time in the design process, for example twice a week for six weeks, participating in activities similar to informants or design partners. In Children as Software Designers [10, 11, 12], children become software designers and developers; adults are not involved in the process other than to teach children the technological skills they need. While there are many possible methods of working with children in the technology design process, over the past fourteen years we have continually developed and employed a design partnering method called *Cooperative Inquiry*.

2.2 Cooperative Inquiry

Cooperative Inquiry is a method of partnering to design technology for children with children [1, 3]. Cooperative Inquiry is a dynamic method which continues to be refined today. Children participate in 90-minute Cooperative Inquiry design sessions at our lab twice a week after school during the school year, and for two weeks of full-day sessions in the summer in a "camp"-type set up. Children on our design team are typically 7 to 11 years old, although Cooperative Inquiry activities have been done with younger children ages 5-6 [4, 13] as well as with pre-teenagers [14].

Children involved in Cooperative Inquiry participate with adults in a wide variety of activities to design technology. The ratio of adults to children is quite high; generally a Cooperative Inquiry design session includes approximately 4-5 adults and 7-10 children. Not only is this a high ratio of adults to children, but it is a small group overall. We find that the high ratio of adults to children, coupled with the small size of the group overall, creates a feeling of a team. We become a community that works together, where no one person, adult or child, has more power than another.

Included within the Cooperative Inquiry method are a variety of techniques. Which technique is used in a design session depends on the current state of a project. These techniques include using low-tech art supplies to create early prototypes, using sticky notes to critique technologies at many levels of development, journaling, and mixing ideas together. Potentially more important than the specific techniques, however, are the overall principles of Cooperative Inquiry.

In Cooperative Inquiry, no one party is considered to be more important than another. We acknowledge each other's strengths. We know that the computer scientist is an expert in programming apps, but that the child is the expert in how 8-year-olds are using iTouches in the classroom. The educator may have a strong theoretical knowledge in the social development of a 10-year-old, but only that child knows how it feels to be the only kid in her fourth grade class whose parents will not let her have e-mail. We respect each other's input and solve problems together. We work on our communication and collaboration constantly, as is necessary for a team which is not only interdisciplinary but also intergenerational. We believe that these underlying principles of Cooperative Inquiry will address some of the needs of tomorrow's education and translate well into the educational setting of the future.

3 The Potential for Cooperative Inquiry in Education

3.1 Motivation

Educators today are charged to prepare children for a world whose future is inherently unpredictable and whose demands are complex and intensified. We no longer need to churn out students who have a basic ability to read, write, and do arithmetic in order to function in menial jobs. Rather, the children of today — who are the leaders and workers of tomorrow — will need to think critically and solve problems ingeniously. They will need to be comfortable with collaborating and communicating not only with co-workers in an office, but also around the world. We suggest one way to help address these issues is to begin to employ the use of methods such as Cooperative Inquiry in education. By doing so, we will value children's voices in their own education and encourage children to become communicators, collaborators, and problem-solvers. Having worked with children design partners for over 15 years [1, 2, 3, 4, 5, 7, 15], we believe that now is the time to move design partnering methods to more formal educational settings.

Recent research [2] has demonstrated that children involved in Cooperative Inquiry can have positive experiences in the social and cognitive developmental areas, including in relationships, confidence, enjoyment, communication, collaboration, skills, and content. Given the positive outcomes of long-term design partners, design partnering could be applied in an educational setting. Schools from pre-kindergarten to grade twelve and beyond could include teaching Cooperative Inquiry not only as a process, but also using it as a method of instruction and experiences. Employing a method such as Cooperative Inquiry in an educational setting encourages working collaboratively to construct knowledge [16]. Vygotsky's work implies working together as a way to support children in learning [17, 18]. This work was furthered with the notion of cognition as a collaborative process [19]. We believe the extension of Cooperative Inquiry into formal classrooms has support from these works.

3.2 Teacher and Professional Training

When implemented in an educational setting, Cooperative Inquiry would involve adults who are less teachers in a traditional sense and more facilitators or collaborators with children in education. In this role, adults would be co-learners with the children. This posits adults and children as partners, which would disrupt the power structure of traditional education environments where the teacher is the primary power figure and children are to be instructed and directed. As this is a large paradigm shift in the role of the adult in education, training and education of adults involved in this Cooperative Inquiry education will be necessary.

As with any method, educators would need to be trained on how to use Cooperative Inquiry in order to implement it as a part of a curriculum. They would need to understand the ways in which this method can be used to encourage students to explore and engage in learning. This training could be accomplished through in-service training for current educators. Courses on the use of technology design methods for children are already in place at major conferences in the field, such as Computer Human Interaction (CHI) [20] and Interaction Design and Children (IDC) [21]. As these conferences are international, their locations change yearly. This could encourage educators world-wide to attend a course in which they could learn about the possible experiences for their students using Cooperative Inquiry.

Training for college and university students studying to become teachers is another option for disseminating information on design methods such as Cooperative Inquiry. As many university programs now include courses on using technology in the classroom, this would be a logical place to insert content on designing technology with children, and the ways in which methods such as Cooperative Inquiry can be implemented in a classroom setting.

3.3 Possibilities for Implementation

Children involved in a Cooperative Inquiry method of education would become not only investigators, but also researchers; and not only learners, but creators of their own knowledge. In order to follow the Cooperative Inquiry model of interdisciplinary team members, adults in Cooperative Inquiry education would be teachers, researchers, professors, professionals from the community, and adult students. Conveying content using Cooperative Inquiry would require the adults involved to ensure they incorporated content and curricula that were appropriate for the child, given the child's developmental needs and interests. However, the child would participate in defining that content.

We are aware of the limitations of large class sizes and adult to child ratio, prescriptive curricula, and standardized testing that persist in public schools, especially in the United States, today. As such, we believe that initial forays into employing Cooperative Inquiry in schools may be best suited to unique situations in public schools, or schools which have lower adult to child ratios and more freedom in curriculum, such as private or charter schools. While our ideal scenario would be children involved in an entire school that employs Cooperative Inquiry, we realize that it might be wise to begin with forays into smaller, more manageable scenarios.

Design partnering for education could be applied to diverse situations which exist in today's schools. Cooperative Inquiry may have potential to be included in classrooms currently configured as classes for students with *special needs*. *After-school programs* are another unique situation in which Cooperative Inquiry could be employed. Aside from public schools, there are possibilities for entire schools to employ Cooperative Inquiry to teaching. These include *private, charter, and technology magnet schools*. We will now explore and explain each of these possible implementations in turn.

Special Education Classrooms. One unique situation within public schools where Cooperative Inquiry may have potential to be included, in classrooms as they are currently configured, is in special education classes. Special education classes are set up to teach children who have special needs, including social, emotional, behavioral, and/or cognitive disorders. Many times, special education includes not only traditional academic learning but also explicit instruction in social and behavioral issues. Class sizes in special education are generally smaller, thus the adult to child ratio is higher, which is a requirement for Cooperative Inquiry.

The social experiences of children on a Cooperative Inquiry design team indicate that these activities could provide positive experiences to children who have social issues. We know that the children who participate in Kidsteam experience positive relationships and confidence as design partners [2]. This was also conjectured by researchers who explored using design partnering with children with special needs [22, 23]. The higher adult to child ratio already in special education classes, coupled with the experiences design partnering fosters in relationships and confidence and the enjoyment children can experience on a design team [2], could prove to be a valuable combination for engaging children with social challenges. This is not to say that a design partnering model should take over a special education classroom, but rather that a teacher, along with adult aides, could choose to employ a design partnering model for selected parts of the curriculum.

After-School Programs. After-school programs are another unique situation in which Cooperative Inquiry could be employed. Children participate in many activities after school, from sports such as gymnastics to art or music classes. Employing Cooperative Inquiry for an after school program could provide children the social and cognitive experiences of communication and collaboration. Instead of a debate club or being on a soccer team, children could be offered the option of a technology design club after school. Although not a formal educational experience, we believe that Cooperative Inquiry could also be an appropriate method in an informal gathering. This club could be limited in size, thus allowing for a lower ratio of adults to children. The club could endeavor to create technology that would in some way benefit their school, such as to solve the problem of too much noise in the cafeteria. This group would experience the communication and collaboration that are inherent in Cooperative Inquiry.

Schools. While it may not be possible to widely implement Cooperative Inquiry into an entire country's public education system, there are opportunities for using Cooperative Inquiry on a school-wide basis in non-public settings. Progressive schools not solely responsible to the public system might have a greater ability to implement design partnering as a mode of education. In the United States, these would include private schools, which are schools funded by tuition and donations and to which parents can choose to send their children if they have the means to do so. Charter schools operate within the public school system but have special authorization via a charter or document that outlines a specific mission for that school. They often have more freedom in how they teach. Finally, magnet schools are schools within the public school system which draw students with specific interest together to a school.

Although Cooperative Inquiry is not intended as a method of teaching and learning in the traditional sense, given the problem solving and spontaneous concept learning experienced by the design partners, there is the possibility that a modified type of design partnering could be used in a formal educational setting. For example, it would be interesting to see if a small classroom of third graders could work together to design a technology to teach other children a specific topic in science, and if through this activity, they experienced science and/or skill learning. Kafai's work with Children as Software Designers [10, 11, 12] indicated that children can learn science, technology and math content through programming software. Similar content experiences may be available to children working on Cooperative Inquiry design teams [2].

An even more in-depth way to apply Cooperative Inquiry is to establish charter schools in technology design. Tomorrow's economy will demand many workers who are skilled in technology design. Since we know from the research cited above that children gain experience in cognitive skills and content as a result of being a part of a Cooperative Design team, a charter school which utilizes Cooperative Inquiry as a significant part of the method of instruction could be established. Such a charter school would include teaching Cooperative Inquiry not only as a process, but also using it as a method of instruction and experiences.

Technology magnet schools are another context in which Cooperative Inquiry could be employed. An important part of the curriculum of these schools would be designing technology. If a high school student is interested in a career in technology design, she should be introduced to a wide variety of design methods early. Not only should there be a broad teaching of various design methods, but students in technology magnet schools should experience working with different design methods to solve the real world problem of designing a technology. This would involve scaling up the model of Cooperative Inquiry to students at a middle school and high school level, ages 12 through 18. This scaling up to older children began with middle school students [14]. In research being conducted today at the University of Maryland, we are investigating not only using Cooperative Inquiry methods with middle school age students, but with high school students as well. Though no formal evaluation has been done to date, we are experiencing early preliminary success in scaling Cooperative Inquiry techniques for use with teenagers.

3.4 Real World Connections

A teacher interested in having her students experience communication and collaboration could employ Cooperative Inquiry as a method for a school project. If an educator were interested in conveying content using Cooperative Inquiry, the key would be to ensure that the technology they were designing incorporated the content that was part of the curriculum of the school. For instance, if a second grade class was learning about their home state in social studies, they could be asked to create a website to teach other children about specific aspects of their state. The results from our work indicate that the experience of working with outside professional partners is powerful to the child design partners as it provides a sense of relevance, illustrates the broad impact and value of the project, as well as models how collaborative relationships work. Educators should consider collaborating with outside professional partners in order to magnify the importance a Cooperative Inquiry project undertaken in the classroom. For example, the second grade class working to create a website about their home state might partner with the state government in order to deploy the technology broadly

In any of these situations, the educator would need to be in a situation in which the adult to child ratio is higher than a typical classroom. Educators may consider asking parent or family members to volunteer in the classroom to help with Cooperative Inquiry activities. Another possibility would be to involve people who are learning to become teachers, including university students enrolled in education programs, in Cooperative Inquiry activities, as well as local experts including those from industry, small businesses, and academia.

3.5 Assessment

Appropriate means of assessment or evaluation is a large field with continual debate within education. There is a large body of research with regards to assessment. An analysis of the many ways to assess children educationally is beyond the scope of this work. Future practitioners and researchers who choose to implement Cooperative Inquiry as a method for formal education would need to adopt, define, explain, and defend a theoretical framework for any quantitative assessments used.

If educators were interested in understanding how Cooperative Inquiry could be employed in an educational setting, and the value of doing so, they may first want to study its effectiveness on a smaller scale. Comparative or intervention studies could be developed between classrooms within the same school in which one employs a traditional method of teaching and another employs Cooperative Inquiry. For example, a traditional classroom and a Cooperative Inquiry classroom of third graders could both spend two weeks studying oceans. In the traditional classrooms, activities such as reading for information, watching videos, and writing reports might occur. Perhaps an oceanographer would come to this classroom and give a presentation about oceans. The Cooperative Inquiry class might spend time collaboratively developing a website to teach other children about the oceans.

In addition to the difference in activities that these classrooms would undertake, the classrooms would have to be administered differently from the initiation of the study. The traditional classroom would have the traditional model of one or two authoritative adults and a group of children approximating an average classroom size, from twenty to thirty students. The Cooperative Inquiry class would have a smaller class size and a higher ratio of adults to children. These adults could be teachers, researchers, and adult students. Perhaps an oceanographer would come to work with this class from time to time, not only to give presentations, but also to work with the students in the development of the website. The Cooperative Inquiry class would need to spend time building the team of adults and students before the intervention took place.

The pre- and post-tests generally administered by teachers to ascertain the content knowledge growth could be used to determine the comparative content learning of those in the traditional versus the Cooperative Inquiry classroom. Studies such as this would provide information on the value of Cooperative Inquiry as an educational mechanism. It would also be important to assess, formally or informally, the more "intangible" learning of the students – were there gains in communication and collaboration in either room? How do these compare? Do we consider enjoyment and confidence important skills to encourage in students? We know that these are experienced by children participating in Cooperative Inquiry, but how do they compare to those in a traditional classroom? These are important questions not only about the value of Cooperative Inquiry in education, but about what we value overall in education.

4 Conclusions

We value children's abilities to collaborate, communicate, and be confident. We believe that these qualities will lead to stronger, more effective, more productive adults in society. Tomorrow's workers will need to be able to communicate, collaborate, and problem solve with partners across the globe. To prepare them new methods of education must be employed. It is our belief that Cooperative Inquiry should be one of these methods. To that end, we believe that Cooperative Inquiry needs to move from a method only used for designing technology to a method for education. There are numerous ways that Cooperative Inquiry can be translated into other environments, such as after school clubs, and into formal educational settings, such as special education classrooms and private, charter, or magnet schools. In the future, we intend to begin employing Cooperative Inquiry method in settings such as these. It is our goal to explore better ways to meet the needs and interests of children in schools today.

Acknowledgments. We would like to thank the 40+ children who have been our design partners over the years. We also thank our adult design partners for their continued work. Finally, we thank David Cavallo, Tammy Clegg, Jason Yip, and June Ahn for pushing the boundaries in thinking about the future of education and for thoughts on this work.

References

1. Druin, A. The role of children in the design of new technology. Behaviour and Information Technology, 21(1), 1-25 (2002)
2. Guha, M. L. Understanding the Social and Cognitive Experiences of Children Involved in Technology Design Processes. Dissertation, University of Maryland (2010)
3. Druin, A. Cooperative Inquiry: Developing New Technologies for Children with Children. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: The CHI is the Limit, pp. 592-599 (1999)

4. Guha, M.L., Druin, A., Chipman, G., Fails, J. A., Simms, S., Farber, A. Mixing Ideas: A New Technique for Working with Young Children as Design Partners. In Proceedings of Interaction Design and Children: Building a Community, pp. 35 - 42 (2004)
5. Fails, J. A., Druin, A., Guha, M. L., Chipman, G. Simms, S., Churaman, W. Child's play: A comparison of desktop and physical interactive environments. In Proceedings of Interaction Design and Children: Toward a More Expansive View of Technology and Children's Activities, pp. 48-55 (2005)
6. Scaife, M., Rogers, Y., Aldrich, F., Davies, M. Designing for or Designing With? Informant Design for Interactive Learning Environments. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Looking to the Future, pp. 343-350 (1997)
7. Chipman, G., Druin, A., Beer, D., Fails, J.A., Guha, M. L., Simms, S. A Case Study of Tangible Flags: A Collaborative Technology to Enhance Field Trips. In Proceedings of Interaction Design and Children, pp. 1 - 8 (2006)
8. Large, A., Bowler, L., Beheshti, J. & Nesset, V. Creating web portals with children as designers: Bonded design and the zone of proximal development. McGill Journal of Education 42(1), pp. 61-82 (2007)
9. Large, A., Nesset, V., Beheshti, J. & Bowler, L. "Bonded Design": A novel approach to intergenerational information technology design. *Library & Information Science Research,* 28, pp. 64-82 (2006)
10. Kafai, Y.B. Software by Kids for Kids. Communications of the ACM, 39(4), 38--39 (1986)
11. Kafai, Y.B. Children as Designers, Testers, and Evaluators of Educational Software. In Druin, A. (ed.) The Design of Children's Technology, pp. 123 - 146. Morgan Kauffman, San Francisco (1999)
12. Kafai, Y. B. Children Designing Software for Children - What can we Learn? In Proceedings of Interaction Design and Children: Small Users - Big Ideas, pp. 11 - 12. (2003)
13. Farber, A., Druin, A., Chipman, G., Julian, D., Somashekhar, S. How Young Can our Design Partners Be? In Proceedings of the Participatory Design Conference, pp. 127 - 131 (2002)
14. Knudtzon, K, Druin, A., Kaplan, N., Summers, K., Chisik, Y., Kulkarni, R, et al. Starting an Intergenerational Technology Design Team: A Case Study. In Proceedings of Interaction Design and Children: Small Users - Big Ideas, pp. 51 - 58 (2003)
15. Fails, J.A., Druin, A., Guha, M. L. Mobile Collaboration: Collaboratively Reading and Creating Children's Stories on Mobile Devices. In Proceedings of Interaction Design and children pp. 20-29 (2010)
16. Papert, S., and Harel, I. Constructionism: Research reports and essays, 1985-1990. Norwood, New Jersey: Ablex Publishing Corporation. (1991)
17. Vygotsky, L. S. Mind in Society. Harvard University Press, Massachusetts (1978)
18. Vygotsky, L. S. Thought and Language. MIT Press, Massachusetts (1986)
19. Rogoff, B. Cognition as a Collaborative Process. In Kuhn, D. & Siegler, R. S. (eds.) Handbook of Child Psychology, Volume 2: Cognition, Perception, and Language, Fifth edition. John Wiley & Sons, New York (1998)
20. Druin, A., Guha, M.L., Fails, J. A. New Methods for Designing with and for the iChild: Strategies for Today's Mobile, Social, and Internet Technologies. Course Presented at Computer Human Interaction, Atlanta, Georgia (2010)

21. Druin, A. Farber, A., Guha M. L. Methods for Partnering with Children to Develop New Technologies. Tutorial presented at Interaction Design and Children, Preston, England (2003)
22. Gibson, L., Gregor, P., Milne, S. Case Study: Designing with 'Difficult' Children. In Proceedings of the International Workshop "Interaction Design and Children", pp. 42 - 52. (2002)
23. Jones, C., McIver, L., Gibson, L., Gregor, P. Experiences Obtained from Designing with Children. In Proceedings of Interaction Design and Children: Small Users - Big Ideas, pp. 69 - 74. (2003)

Effects of Print-Storybooks and E-Storybooks with Reading Comprehension Strategies on Fifth Graders' Reading Comprehension Ability

Hsiu-Shuang Huang[1], Shang-Liang Chen[2], Yea-Mei Leou [3], Ho-Chuan Huang[4], Ching-Yu Yeh[1], Yun-Yao Chen[2], Chun-Lien Chen[1], and Ya-Ying Tseng[5]

[1] National University of Tainan, Department of Education, 33, Sec. 2, Shu-Lin St., Tainan 700, Taiwan, R.O.C.
hshu@mail.nutn.edu.tw
[2] National Cheng Kung University, Institute of Manufacturing Information and System Affilliation, 1.Ta-HsueH Road 701, Tainan, Taiwan, R.O.C.
slchen@mail.ncku.edu.tw
[3] National Tainan Institute of Nursing, 78, Sec. 2, Minzu Rd. 700, Tainan, Taiwan, R.O.C.
ymleou@mail.ntin.edu.tw
[4] National Kaohsiung University of Applied Sciences, Departmaent of Information Management, Chien Kung Campus, 415 Chien Kung Road, Kaohsiung 807, Taiwan, R.O.C.
hchuang@cc.kuas.edu.tw
5National Changhua University, Department of Education, 1, Jin-De Road, Changhua, Taiwan, R.O.C.
fang9487@ms1.hinet.net

Abstract. The goal of the study was to develop the instruction module of the e-storybooks reading comprehension in the elementary school and test the effect of the instruction module of the e-storybooks reading comprehension. This study was to compare the influences of print-storybooks and the instruction module of the e-storybooks reading comprehension among the fifth graders' reading comprehension ability. The results of the study indicated that there was significant difference between the experimental group and the controlled group. The controlled group had better performances of reading comprehension than experimental group. To put it more specifically, students and teachers were busy learning how to operate the module and platform, and that resulted in greatly reduced learning effect. However, with the mastery of e-storybooks instruction module, experimental group class expressed higher motivation than controlled group class in learning activities. These data indicated that the study had the positive contribution.

Keywords: digital reading, e-storybooks, reading comprehension, reading strategies.

1 Introduction

Because the rapid improvement of technology in this e-age, the types of reading change a lot. Traditionally, people read printed books. Nowadays, people try e-books

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 570–579, 2011.

and e-reading besides reading books. E-book reading seems to become a new trend in the future. Learners have become more actively involved in searching for information and reading on the web because internet learning had been transformed from one-way communication into interactive communication.

According to the report of PISA 2006, the Taiwan students' math mean score was the highest, science mean score was the fourth, but the mean score of reading was the sixteenth among the 57 countries that participated in the PISA. It showed that the Taiwan students were weak in reading literacy and lacked abilities of reflection and evaluation. The results made us reconsider our education and pay more attention to the reading instruction. Reading was the core of learning and should be emphasized[1][7].

However, it was a pity that e-book and Chinese reading machine were still in the initial stage in Taiwan, and the environment of digital reading was not matured enough. What students read on line is "shallow reading' which was short and can be read fast, such as e-newspapers[2]. Nevertheless, developing reading comprehension strategies of reading e-books in order to deepen students' meta-cognition is very important.

Therefore, the researchers hoped to base on the electronic storybooks and develop suitable e-storybooks integrating with reading comprehension strategies. And then, the researchers attempted to compare the efficiency of different groups with reading print-storybooks and e-storybooks . The purposes of this study were as follows:

1. To develop instruction modules of e-book reading comprehension in order to enhance the readers' reading strategies and reading comprehension ability.
2. To explore the changes of the fifth graders' reading comprehension in the instruction module of print-storybooks and the instruction module of the e-storybooks.

2 Literature Review

2.1 Digital Reading

Digital reading included reading digital materials on-line or off-line. Digital reading could not only help people read a lot of data but also transfer people's reading habits. The characteristics of digital reading included: (1) being able to present context with multimedia. (2) providing information quickly. (3) helping leaping reading. When doing traditional paper reading, readers often followed the authors editing formats. However, with the functions of hyperlink of digital reading, readers could choose any contents to read instead of reading from the start to the end. Reading could be done more flexibly. Readers could get the important point of the content in a short time. (4) being helpful to reading and writing, because it could offer opportunities of the interactions between writers and readers. (5) reading whenever and wherever, as long as the readers have computers.

In short, the media of paper reading or digital reading varied the reading context. Print-books had clear boarders and fixed contexts. However, the borders of the digital reading were blurred with multiple contexts. Although digital books were opener than print-books, readers could take the same reading strategies to enhance the reading comprehension.

2.2 E-Storybooks

E-storybook, which combined with picture books, fairy tales and multimedia, is one kind of e-books. There are four characteristics about e-book. Firstly, e-books could be downloaded so that it could help readers get instant and flexible information. Secondly, e-books could provide supportive information through the computer systems to overcome the problems of learners' shortage of background knowledge. Thirdly, the limits of space and time would not affect the feedback system of e-books. Fourthly, e-learning needed to combine related techniques with computer systems[2]. Leu did a study of e-books experimental teaching on the students. The results showed that the students in the experimental group could have more detailed descriptions than the students in the controlled group because of the vivid images of e-books[3].

2.3 Reading Comprehension in Digital Reading

On-line reading comprehension allowed readers to make use of internet and other communication skills to find out important questions, search for information, analyze useful data, and communicate [4].

Coiro and Dobler did a study on the expert readers of the sixth graders and limited them to find and evaluate information. The results showed that there were similarities between print reading and on-line reading but the on-line reading was more complicated[5]. However, there is no research being done in e-storybooks, which were important and useful to learning and teaching. Therefore, the focus of this study was to compare the influences of print-storybooks and e-storybooks with reading comprehension strategies on the fifth graders' reading comprehension ability.

Thus the researchers followed Elizabeth's viewpoints to integrate four reading strategies into the instruction module of the e-storybooks reading comprehension, which included "prediction", "inference", "query" and "summary". The main reading strategies of the instruction module illustrated in Table 1.

Table 1. Reading strategies of the instruction module

Reading strategies	Content
prediction	Readers could predict the content of e-books according to their titles, clues, outlines and pictures.
query	Readers could identify the main messages in e-books and ask related questions. In other words, readers could monitor the reading comprehension process by asking themselves questions.
inference	Readers could integrate their prior knowledge with textual clues to comprehend the core context.
summary	Readers could retell core elements in the stories with their own words.

As mentioned above, the instruction module of the e-storybooks reading comprehension needed digital interactive scripts to arrange presentation of media, interactive links and feedback. In order to constructive reading recycle made the reader through decoding, meaning comprehension, inferential comprehension and monitoring comprehension, the instruction module adapted the design of scaffolding. In the instruction module of the e-storybooks reading comprehension, reader could read the multimedia e-storybooks with four reading strategies, every strategy was displayed in four steps: "test and feedback", "inform strategy", "practice to use" and "online discussion and clarification". Through the gradual reduction of the assistance, repetitive practice and immediate assessment, the readers gradually developed the capacity of self-learning and independence. The structure of the instruction module showed in Fig. 1.

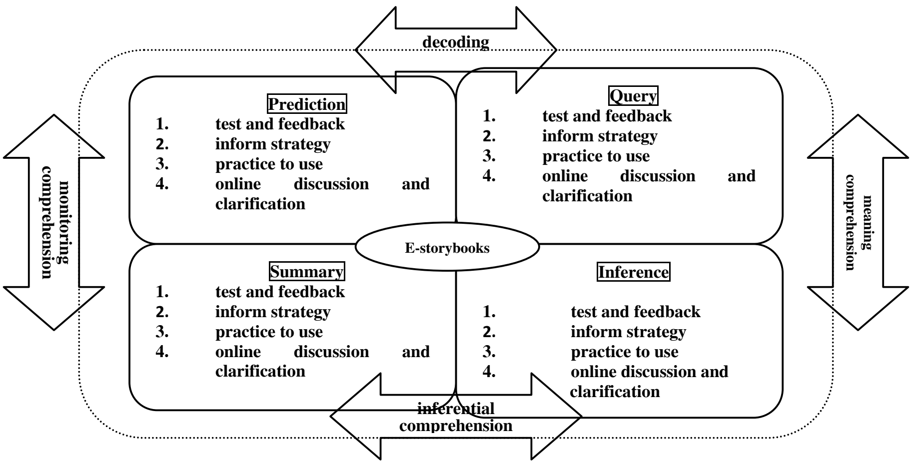


Fig. 1. The structure of the instruction module

3 Method

3.1 Subjects

Two classes of fifth graders were participated in a pretest and posttest quasi-experimental design. There were 31 students in the experimental group class, which had an average age of 11.5, with 17 girls and 14 boys. The other 32 students were in the controlled group class, which had an average age of 11.3, with 17 girls and 15 boys.

3.2 Process

Before the intervention started, the researchers analyzed different Chinese storybooks and discussed with some experienced elementary school teachers to make good choices. Then, the researchers created the e-storybooks instruction module and E-Book Reading Platform. The instruction module of the e-storybooks reading comprehension had the same contents as the print-storybooks, but they were presented in the interactive educational activities and extended learning field by E-Book Reading Platform. The experimental group class went through e-book reading with instruction module of the e-storybooks reading comprehension for 4weeks, 80minutes/week. The controlled group class went through Print-storybooks reading

Table 2. The design of the study

Group	N	pre-test	experimental intervention	post-test
The experimental group class	**31**	Y_1	X_1	Y_2
The controlled group class	**32**	Y_3	X_2	Y_4

Y_1, Y_3 : took pre-test of reading comprehension test.
Y_2, Y_4 : took post-test of reading comprehension test.
X_1 : E-storybooks reading with instruction module of the e-storybooks reading comprehension for 4 weeks, 80minutes/week.
X_2 : Print-storybooks reading with reading comprehension strategies for 4 weeks, 80minutes/week.

with reading comprehension strategies for 4 weeks, 80 minutes/week. All the students took pre-test and post-test of reading comprehension tests before and after the intervention. The design of the study is displayed in Table 2.

3.3 Design of E-Storybooks Reading Platform

The system set up the server environment in Xoops 2.2.4 and adopted APHP, HTML, Flash and other technologies to establish the learning platform. Teachers provided reading materials and issues in the database, so that students could download e-storybooks through internet.

Before using the platform, firstly, each participant should set up a password and a username. In e-storybooks reading sub-system, registers downloaded the instruction module of the e-storybooks reading comprehension and followed the interactive interface to participate reading activity. In discussion and feedback sub-system, teachers and students could enter the web-based forum of platform and exchanged opinions. Additionally, students followed the procedures to answer questions, then e-storybooks reading platform provided the pre-set feedback. In e-storybooks management sub-system, teachers managed e-storybooks and the learning process records. The functions of the e-storybooks reading platform were implemented in Fig. 2.

In Fig. 2, teachers expanded e-storybooks instruction module through e-storybooks management sub-system, and guided the discussion through discussion and feedback sub-system. Besides, students could read e-storybooks on-line or off-line. Teachers and students could enter the web-based forum of platform to clarification important issues. From the above, the interface of e-storybooks reading platform showed in Fig. 3.

In Fig. 3, teacher provided e-dictionaries, search engines and relevant websites on the web, students could search information to clarify the confusion and supply the lack of prior knowledge. Additionally, students could overview the relevant websites in order to reduce the distraction. As the result, the platform could make up for the lack of interactions between learners and teachers in e-storybooks learning environment.

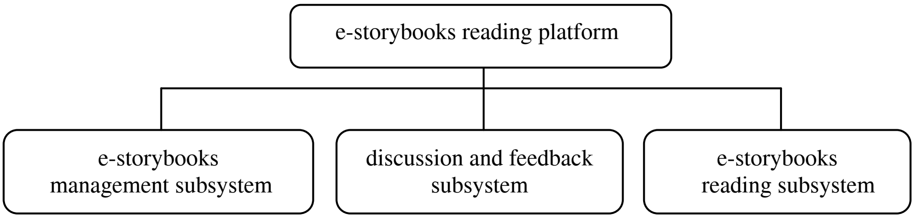


Fig. 2. The functions of e-storybooks reading platform

Fig. 3. The interface of e-storybooks reading platform

3.4 Instruction Module of the E-Storybooks Reading Comprehension

In this study, the instruction module focused on how to enhance reading comprehension ability through interactive e-storybooks. The interactive e-storybooks played a virtual teacher to guide reader and avoid students from losing in multimedia hyperlinks. The interactive e-storybooks were divided into four parts to foster reading comprehension strategies. Every strategy was displayed in four steps: “test and feedback”, “inform strategy”, “practice to use” and “online discussion and clarification”. The design of the instruction module showed in Table 3.

3.5 Design of the E-Storybooks

According to the theory of reading comprehensive strategies, researchers applied ADDIE Model to develop the interactive e-storybooks. The interactive e-storybooks integrated the content of storybooks and educational activities. In order to use the characteristic functions of interactive multimedia, we developed the more interesting and vivid interface to attract attention from users and to promote their reading comprehension ability. The e-storybooks materials were made by Adobe Flash and Action Script2.0 syntax. The design of the interactive e-storybooks materials was showed in Fig. 3, which were divided into five aspects: the context, prediction activity, query activity, inference activity, and summary activity. The interactive type of reading activities included object dragging, interactive buttons, hyperlink pages, animation and painting. The e-storybooks reading followed the text structure to present activities and to provide questions that guided students to discuss the contents.

Table 3. The design of the instruction module

Strategy Sub-module	Instruction Sub-module	Content
Prediction	test and feedback	■ Testing the ability of prediction ■ Creating cognitive conflicts ■ Formatting learning hypotheses
	inform strategy: title page practice to use : picture clue	■ Recognizing the principles of prediction ■ Practicing how to predict through pictures and flashes
	online discussion and clarification	■ Presenting the thought of prediction ■ Admiring the classmates and asking questions
Query	test and feedback	■ Testing the ability of asking questions ■ Creating cognitive conflicts ■ Formatting learning hypotheses
	inform strategy : 6W	■ Recognizing the principles of asking questions
	practice to use	■ Practicing how to asking questions in e-book reading
	online discussion and clarification	■ Connecting to web-based forums ■ Designing the questions based on the context
Inference	test and feedback	■ Testing the ability of inference ■ Creating cognitive conflicts ■ Formatting learning hypotheses
	inform strategy : context clue/ prior knowledge	■ Recognizing the principles of inference
	practice to use	■ Combining life experiences with context clue trough multimedia flash games ■ Exploring the opinions of e-storybooks
	online discussion and clarification	■ Connecting to web-based forums ■ Presenting the viewpoints of e-storybooks
Summary	test and feedback	■ Understanding the plot through flash game
	inform strategy : story structure	■ Presenting systematical structures and tips
	practice to use	■ Readers used the flash paint to mark important sentence. ■ Reader could click on the answer and compare with the teacher's thought
	online discussion and clarification	■ Using cognitive map to organize the important sentences. ■ Connecting to web-based forums ■ Presenting summary

Students not only read e-book but also participated the flash games to check how much they had learned from the e-storybooks. At the same time, teachers enhanced the readers' comprehension and motivation through hyperlinks of context and interactive feedback.

3.6 Measurement and Data Analysis

Researchers used “Reading Comprehension Assessments” to understand the differences before and after the intervention. The assessment tested students’ changes of reading comprehension ability. In order to set a national norm, the participants were 2462 elementary school students in different areas of Taiwan. The assessment consisted of twelve sub-tests which were designed for the second, third, fourth, fifth, and sixth graders. There were two similar texts for each grade. The coefficients of internal consistency were between .70 and .86. The coefficient of test-retest reliability was between .70 and .94. Comparing to other related assessments, the coefficient of criterion-related validity was between .21 and .78. In short, the reading comprehension assessments had good reliability and validity [6]. The collected data were analyzed by Descriptive statistics and an Independent Sample t-test, which was to compare the mean scores of the progress score differences (post-test score – pre-test score) between the experimental group class and the controlled group class.

4 Result

In order to compare the performances of reading comprehension between the experimental group and controlled group, the researchers collected the pre-test scores and post-test scores in Reading Comprehension Assessments to understand the differences, the result indicated in Table 4.

In Table 4, the post-test average of reading comprehension decreased, and the post-test average of experimental group was lower than the controlled group. Then we compared the progress score differences (post-test score – pre-test score) with an Independent Sample *t*-test, the result indicated in Table 5.

Table 4. Descriptive analysis of the performance of reading comprehension

Reading Comprehension Assessmer	Group	*N*	Mean	*SD*
pre-test	the experimental group	31	16.81	3.28
	the controlled group	32	16.25	4.66
post-test	the experimental group	31	16.35	3.50
	the controlled group	32	16.38	4.52

Table 5. Independent Sample *t*-test of the performance of reading comprehension

group	Mean	*N*	*SD*	*df*	*t*	*p*
the experimental group	-13.16	31	6.59	61	-10.90	.000
the controlled group	.13	32	1.64			

In table 5, there were significant differences between the experimental group and the controlled group in the progress score (t=-10.90, p<.05). In other words, the controlled group had better performances of reading comprehension than experimental group.

5 Conclusion

The study was designed to test the effect of the instruction module of the e-storybooks reading comprehension. Based on multimedia electronic-storybooks, the instruction module applied reading comprehension strategies into context, students not only read the multimedia context, but learned reading comprehension strategies. Additionally, the instruction module attempted to transfer personal reading type into cooperative reading type through the platform, on which students could search information and exchange viewpoints in web-based forums.

The result showed that there was significant difference between the experimental group and the controlled group in reading comprehension test. The controlled group had better performance than the experimental group. The result was out of our expectations. In order to realize the users' viewpoint of the instruction module of the e-storybooks reading comprehension, the researchers randomly selected four students from good academic performance group and lower academic performance group to interview. Jane, who had the good academic performance and always completed the learning mission in the study, said: *I did not know why to be so complex. Without these strategies, I could comprehended the meaning.*

The same dilemma was displayed by the teacher. Elaine, who was the teacher in experimental group and controlled group reflected: *I was rush to get familiar with the e-book module, l was worried about that my students did not know how to use e-book to learn, so I ignored to lead children to pay attention to the reading comprehension.*

However, Tom had expressed different opinions, who had the lower academic performance and always spent more time on learning mission of e-storybooks instruction module, said: *I not only clicked buttons of the e-storybooks to acknowledge meanings of words, but read with reading comprehension strategies back and forth. Before adapting e-storybooks instruction module, it was strange for me to read with comprehension strategies. How magic these experiences were!*

In short, the e-storybooks instruction module could response to the need of learner. To sum up, the new way of reading and teaching made students and teachers feel confused. The students and teachers were busy learning how to operate the module and platform, and that resulted in greatly reduced learning effect. However, with the mastery of e-storybooks instruction module, experimental group class expressed higher motivation than controlled group class in learning activity. According to the record of classroom video in the sixth class, the controlled group felt impatient at writing summary , they usually complained during the learning activities. On the contrast, the experimental group class enjoyed in writing summary, exchanging information and clarifying the puzzle through web-based forums. These indicated that the e-storybooks instruction module was helpful to promoting engagement.

As for the limitations of this study, firstly, in this study, the experimental group class went through the instruction module of the e-storybooks reading comprehension only for 4weeks, 80minutes/week. The experimental time was too short to verify the effect of the instruction module of the e-storybooks reading comprehension. Secondly, the teachers and students spent a lot of time to get familiar with the application of the instruction modules, and ignored the reading comprehension ability. Therefore, researchers who were interested in this topic could further design the sub-module to explain the instruction module of the e-book reading comprehension or they could extend experimental time to test effect of the instruction module of the e-storybooks reading comprehension in future studies.

Acknowledgement

The financial support from the National Science Council (NSC99-2420-H-024-002) was highly appreciated.

References

1. The report of Taiwan in PISA (2006), http://www.sec.ntnu.edu.tw/PISA/PISA20062008
2. The study of Digital reading eye movement patterns measurement on development and Application, http://dspace.lib.fcu.edu.tw/handle/2377/28218
3. Leu, D.J., et al.: Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In: Ruddell, R.B., Unrau, N. (eds.) Theoretical Modules and Processes of Reading, 5th edn., pp. 1568–1611. International Reading Association, New York (2004)
4. Elizabeth, S.D.: Reading on the Internet: The link between literacy and technology. Journal of Adolescent & Adult Literacy 47, 80–83 (2003)
5. Coiro, J., Dobler, E.: Exploring the online reading comprehension strategies used by sixth-grade skilled readers, to search for and locate information on the Internet. Reading Research Quarterly 42, 214–257 (2007)
6. Ko, H.W.: Reading Comprehension Test. Psychological Testing 46(2), 19–35 (2000)
7. Leu, D.J.: Expanding the reading literacy framework of PISA, to include online reading comprehension. A working paper commissioned by the PISA Reading Expert Group (2009)

The Interaction of Children's Concepts about Agents and Their Ability to Use an Agent-Based Tutoring System

Alicia M. Hymel[1], Daniel T. Levin[1], Jonathan Barrett[2], Megan Saylor[1], Gautam Biswas[2,3]

[1] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37240, USA
[2] Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37240, USA
[3] Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN 37240, USA
{Alicia.M.Hymel,Daniel.T.Levin,Jonathan.I.Barrett, Megan.Saylor}@vanderbilt.edu, Biswas@eecsmail.vuse.vanderbilt.edu

Abstract. Computer-based teachable agents are a promising compliment to classroom instruction. However, little is known about how children think about these artificial agents. In this study, we investigated children's concepts about the intentionality of a software agent they had interacted with and tested whether these concepts would change in response to exposure to the agent. We also tested whether individual differences in concepts about agent intentionality would affect children's ability to learn from the agent. After repeated exposure to a teachable agent, students did not make more intentional attributions for the agent than a computer, but a general understanding of agency predicted success in learning from the agent. Understanding basic concepts about agency appears to be an important part of the successful design, implementation, and effectiveness of computer-based learning environments.

Keywords: learning, artificial agents, theory of mind.

1 Introduction

Computer-based learning environments have proven to be a valuable resource for both students and educators. Of particular promise are environments featuring teachable agents – graphical representations of characters that students can teach using speech, text, or visual representations (e.g., [1]). A central assumption underlying these systems is that the agents exhibit behaviors that are similar to human characters, which may invoke cognitive processes for learning and monitoring of one's learning more effectively than less social systems. However, little is known about the specific cognitions that underlie this putative benefit, and almost nothing is known about the relationship between students' specific beliefs about these agents and their more general understanding of concepts such as agency. Although students

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 580–589, 2011.

most likely understand that these characters are not actually people and that they do not have human "brains" or a fully functional cognitive system, it is not clear whether students attribute specific cognitive properties to these agents that are similar to those they would attribute to humans. In addition, these agents are clearly designed to invoke broad concepts about intentionality and agency, but an important possibility is that they go beyond invoking these concepts to actually changing them. Conversely, if human-like artificial agents more effectively invoke cognitive and metacognitive processes than agents dissimilar to humans, individual differences in beliefs about artificial agents can affect the use of these systems. Thus, we believe that the link between concepts about agency and utilization of teachable agents is a two-way street, and that concepts about agents will not only affect the use of these systems, but this experience will affect concepts about agents.

To explore children's concepts about teachable agents and the interaction between these concepts and children's use of the systems, we employed a newly developed "behavioral prediction" measure of concepts about agency to understand how students think about a specific teachable agent called Betty. We first measured children's behavioral predictions for multiple agent types to determine whether their beliefs about agency differ from adults. We also tested whether students who used the teachable agent system for extended lessons on climate change and food webs were better able to predict differences in agency between machines and people, and, conversely, whether students who showed a particularly strong understanding were more successful in learning the course content presented in the system. Below, we begin by summarizing basic principles of agency and theory of mind, review a previously utilized measure of intentionality, and discuss reasons why this interplay between concepts about agents and the use of teachable agent systems may occur.

1.1 Basic Concepts About Agency

The body of research on theory of mind development constitutes a good starting point for understanding concepts about software agents. Using theory of mind, individuals are able to comprehend the mental states of other agents, as well as predict their future thoughts, emotions, and actions [2]. These predictions can be guided by how the agent is believed to mentally represent information. Intentional representations are characteristic of human thought, and are closely linked to their referents. One referent cannot be freely substituted for another, as the contents and meaning of that referent are important [3]. Non-intentional representations, on the other hand, are more characteristic of computers. These representations are less closely linked to their referents [4]. By this definition, representations serve as symbolic placeholders that the system acts on with little importance placed on their contents. One way of summarizing this contrast is to suggest that intentional representations reflect truly situated semantic knowledge about the world while non-intentional representations are more like abstract pointers that confer little real knowledge to a representing system that does not really "know" the true meaning of the representations.

Over the first few years of life, children come to understand people's actions as being driven by intentional mental representations that underlie beliefs, desires, and goals. A key component of this understanding is illustrated in a study conducted by Woodward [5]. Infants were shown either a human hand or an inanimate object (e.g., a mechanical claw) moving towards one of a pair of objects. After several repetitions of this event, the locations of the two objects were switched, and the hand or inanimate object either moved toward the previously reached-for object in its new location, or the new object, now in the previously acted-upon location. Woodward hypothesized that the hand that moves towards the previously reached for object in the new location is behaving consistently with intentional thought processes. The action is explainable based on a goal that is supported by an intentional representation of a real-world object, and because this goal has remained consistent across all trials infants should not be surprised to see the hand reach to the previous object even though it is in a new location. Alternatively, the hand that moves towards the new object in the previously used location is behaving consistently with non-intentional thought processes. Rather than acting upon an object, this agent is repeatedly acting on a location, meaning that the goal object can be freely substituted across trials without consequence.

Woodward found that when a human hand was reaching, infants looked longer when it reached towards the new object at the old location. In contrast, when the actions were performed by the stick-like inanimate object, there was no difference in looking time. This finding suggests the infants interpreted the reach by the hand as a goal-driven intentional action on an object, but did not do so for the inanimate stick. We used the underlying logic of this paradigm to construct a questionnaire to investigate adults' understanding of the mental representations and actions of human and non-human agents. Participants were asked to make predictions about the behavior of multiple agent types (e.g., a human, an ambiguous mechanical agent, and a computer) after reading a description of the agent's previous actions. This description was held constant across all agents. Participants then selected one of two actions as the agent's next step, one of which was consistent with intentional reasoning, while the other indicated non-intentional representations.

1.2 Measuring Attributions of Intentionality

In previous experiments using our measure (see Methods section for questionnaire details), we found that adults are able to distinguish goal-directed human representations from non-intentional mechanical representations and believe that ambiguous agents, such as robots, are similar to computers [6]. This is likely influenced by individual differences in beliefs about ambiguous agents, as individuals who believe that non-human agents are good at understanding human goals do not demonstrate as large of a disparity in their behavioral predictions for the two agent types [6]. These results do not seem to be driven by perceived limits in current technology [7] or by perceived intelligence of the agents [6].

Follow-ups to these basic findings have demonstrated that adults' expectations can be modified by experience. For example, when participants were asked to track a robot's focus of attention as it looked at objects and to remember which object the robot preferred, participants believed the robot more likely to engage in intentional thought than a computer [6]. However, simpler manipulations such as giving the robotic agent anthropomorphic labels (e.g., a human name and describing it as having goals) or showing participants a video of the robot did not lead subjects to differentiate it from computers. These results suggest that interactivity and continuous monitoring of a nominal mechanical agent's mental state may begin to generate expectations for intentional representations in these agents.

1.3 Causes and Effects of Intentional Attributions

While previous experience likely influences beliefs about agents, it is not clear whether any given experience will increase or decrease perceived differences among agents. Some findings imply that extended experience with a mechanical agent is associated with attributions of intentionality to that agent, which would lessen differences in attributed intentionality between humans and machines over time. For example, Nass and Moon [8] found that even expert computer users with extensive email and word processing experience act as though computers have inner states and the ability to feel emotional pain. Conversely, experience may increase the apparent difference between humans and machines. Even infants are experienced enough to generate differing expectations about the movement and goals of people and inanimate objects [9, 10] and are able to distinguish between the object-directed actions of a person and location-directed actions of a machine [5].

While experience affects concepts about agents, it is also possible that concepts about agents affect one's experience with them. For example, it is possible that users who are able to fully distinguish teachable agents from real intentional agents have a learning advantage. Such users may have a more mature understanding of theory of mind, and would therefore be more prepared for the ways such artificial agents differ from humans, lessening the likelihood that unexpected differences cause misunderstandings that interfere with learning. Such increased understanding could ensure metacognitive monitoring techniques promoted by a teachable agent are successfully utilized.

1.4 Exploring the Interaction between Concepts and Experience Using a Teachable Agent System

The relationship between concepts and experience with agents is complex and could affect one's ability to successfully use a system featuring an artificial agent, such as a teachable agent system. The ambiguity introduced by working with an agent that neither appears completely human nor completely mechanical may have a particularly strong effect on children, whose theory of mind is not fully developed. In this experiment, we utilized a teachable agent system ("Betty's Brain") to investigate whether kids differentiate agents in a manner similar to adults, whether concepts about an agent affect children's ability to learn from the agent, and whether concepts about artificial agents are affected by extended experience with them. Students interacted with a teachable agent (Betty) through a series of classroom lessons in which they were asked to teach Betty and give her quizzes to assess her understanding. Betty engaged with the students and expressed a desire to learn. We used our previously developed questionnaire, described below, to measure students' ability to predict the mental representations and actions of a human, a computer, and Betty.

If students' experience affects their concepts about agents, this will be reflected in their behavioral prediction questionnaire responses. If experience with an agent is associated with intentional attributions, repeated exposure to the Betty agent should lead to increased attributions of intentionality for Betty, with no changes in attributions for other agents. However, if experience with a mechanical agent helps to differentiate between humans and machines, students may expect Betty and the computer to perform non-intentionally after interacting with Betty. For example, students may discover that the artificial agent is not capable of reasoning in a manner similar to their own. Alternatively they may draw upon the experience to deepen their understanding of the differences between machines and people, and predict both more intentional behaviors for people, and less intentional behaviors for the computer and Betty. Finally, if concepts about agency affect students' ability to learn from the teachable agent, there should be a correlation between behavioral predictions and quality of students' concept maps in the Betty's Brain system.

2 Methods

2.1 Participants

Participants were recruited from five classrooms in a Nashville, Tennessee public middle school. A total of 108 7^{th} graders (57 experimental and 51 control) were enrolled in the study, and 74 students (69%) completed both the pretest and the posttest. Age and sex were not collected from the participants. Informed consent was obtained from all students and at least one legal guardian of each student.

2.2 Materials

Students completed our previously developed behavioral prediction questionnaire in which they were told they would be asked questions about Betty, a computer, and a person. They were given pictures and a short description of each agent. For example, the description of Betty stated: "Betty is a part of the teaching exercise you just completed. Think of how Betty acts and what makes her think." Next, participants made behavioral predictions. The first prediction scenario drew closely from Woodward's infant paradigm. Participants saw an image of a pair of objects on a labeled grid and were instructed to imagine that the agent chose one of the objects (the toy duck at location A-1). The second object (toy truck at location C-3) was not mentioned. They were then prompted to imagine that the agent repeated the action. Finally, subjects were shown an image of the toy duck and toy truck after their locations had been switched, and were asked to choose which of the two objects the agent would select (the toy truck at A-1, or the toy duck at C-3). Following from Woodward's results, if participants believe the agent to be acting in an intentional and goal-directed manner, they should predict that the agent would select the same object, that is, maintain the same goal. However, if participants believe the agent to be acting in a rote or non-intentional manner, they should predict the agent would maintain its movement pattern without regard to goal state by reaching to the new object at the old location. After completing the behavioral prediction for the first agent, participants made predictions for each of the remaining two agents.

Next, participants completed a second set of behavioral predictions. In this scenario, participants saw seven objects arranged in a horizontal line. The first, third, fifth, and sixth objects were writing instruments, while the second, fourth, and seventh were of varying object type (spoon, scissors, and screwdriver). Participants were told that the agent reached towards the first, third, and fifth items, all of which were writing instruments. They were then asked which object the agent would choose next: the writing instrument located in the sixth position, or the non-writing instrument located in the seventh position. Responses were coded as intentional if the participant selected the writing instrument, as it serves a similar purpose, and helps achieve the same goal, as the previously selected objects. Selecting the non-writing instrument in the seventh position was considered non-intentional as while the spatial pattern of reaching was maintained, the object's purpose was different from the previously selected items.

Participants then completed the final set of behavioral predictions. Participants saw a picture of six items and were told that these items were shown to the agent. The objects could be categorized in one of two ways: taxonomically (office supplies and candy), or perceptually (large, dark, rectilinear objects and small lightly-colored objects). Participants were shown each of these two groupings and were asked to select which way the agent would organize the objects. The use of a taxonomic, or category-based, organizational strategy was considered to be representative of intentional thought, as the objects are grouped according to their function and the intent of their creators [11]. Perceptual categorization relying on the examination of surface features was considered non-intentional, as it requires no knowledge of object meaning or goal states.

2.3 Procedure

Students were assigned by classroom into either the experimental or control condition. Students in the experimental condition used Betty's Brain to complete lessons, and students in the control group completed traditional classroom assignments to learn the same material. Students in both groups learned about arctic climate change and food webs.

Betty's Brain is an agent-based learning environment in which students create causal concept maps to teach Betty, an interactive agent represented by an animated face. The software was designed to promote and reinforce metacognitive techniques, such as knowledge state monitoring, as students must ensure that Betty understands the material sufficiently for her to perform well on quizzes. Students use the Betty's Brain program by reading provided resources and identifying the causal relationships present among concepts described in the text. Concepts and their causal relationships are later entered by each student into a concept map. Students are able to ask Betty questions about the concepts or direct her to take quizzes from a mentor agent (Mr. Davis) to assess her learning. Betty's answers are always logically drawn from the student's concept map and may be incorrect if the concept map is erroneous or incomplete. An example of the Betty's Brain user interface is provided in Figure 1.

Betty is programmed to interact with students as they construct their concept maps. Betty encourages students to read the resources and learn new information so they can teach it to her. She is able to initiate conversations by restating recently taught knowledge and its effects on established causal chains. Betty also requests that students ask her questions to ensure she understands new causal relations. Students

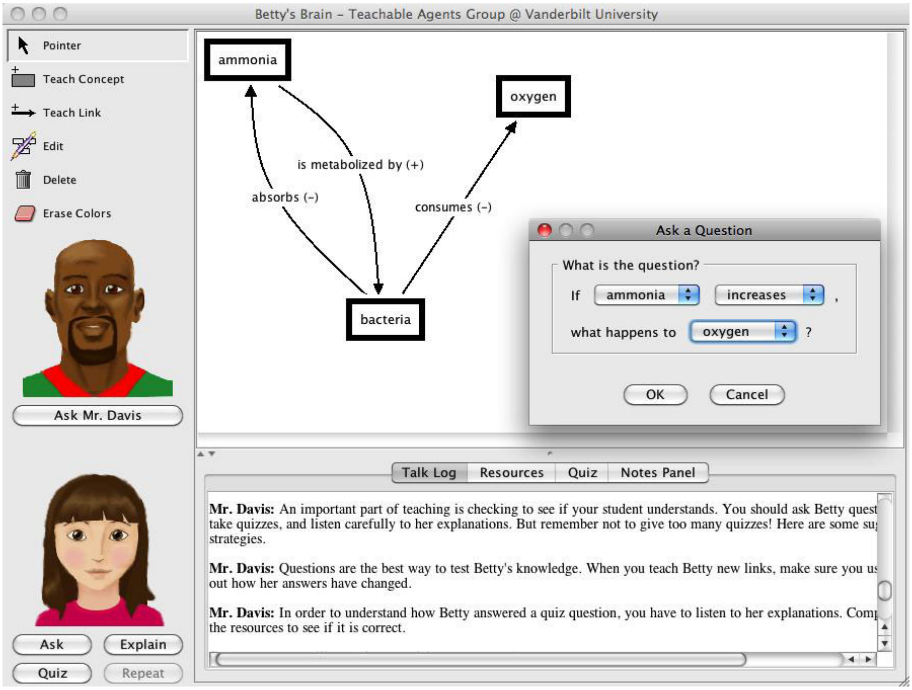


Fig. 1. The Betty's Brain user interface, featuring the Betty and Mr. Davis agents, student-created causal concept map, pop-up window used to ask Betty questions, and sample conversation with Mr. Davis

can ask Betty to explain the logic behind her answers, and she responds using speech, animation, and text. Betty expresses desire to improve her scores on quizzes and disappointment if this goal is not met.

The lessons were split into two units: arctic climate change and arctic food webs. Students in both conditions first took a pretest to establish their baseline knowledge of arctic climate change. After a brief introduction to the arctic climate, students in the experimental condition underwent one class period of training in the Betty's Brain program, while students in the control condition continued with normal lessons. Experimental students then spent four full class periods constructing their concept maps and teaching Betty. After completion, students took a post-test identical to the pre-test. Both groups of students then repeated the series of activities for the arctic food web lessons.

After completing the learning sessions and the second post-test, students were given the behavioral prediction questionnaire asking them to make predictions about a human, a computer, and Betty. Control participants with no previous exposure to Betty were given a brief description of the teachable agent before completing the questionnaire.

3 Results

Students' responses to the behavioral prediction scenarios were very similar to those given by adults [6]. They made similar proportions of intentional predictions for the anthropomorphic agent (Betty; 41% intentional responses) and the computer (43% intentional responses) and gave more intentional responses to the human agent (63%) than for either Betty ($t(72)=4.437$, $p<.001$), or the computer ($t(72)=3.687$, $p<.001$).

We also tested whether behavioral predictions were related to better performance on the learning task, as measured by the quality of the concept maps that the students created. Intentional behavioral predictions for humans and pretest scores were entered into a regression model predicting the average number of correct concepts and links in the concept maps that students created to teach Betty. The overall regression model was significant ($F(2,38)=8.59$, $p=.001$) and accounted for 27.5% of the variance in concept map quality. Both behavioral predictions and pretest scores contributed significantly to the model (standardized $\beta=.278$, $t(37)=2.05$, $p=.048$; and standardized $\beta=3.301$, $t(37)=.45$, $p=.001$, respectively). Thus, an advanced understanding of human agents was associated with increased learning from Betty's Brain.

Two similar regression models were created by substituting the human behavioral predictions for behavioral predictions for the computer and for Betty. While both models were significant ($F(2,38)=5.87$, $p=.006$ and $F(2,38)=5.86$, $p=.006$, respectively) neither the behavioral predictions for the computer (standardized $\beta=-.021$, $t(37)=-.145$, $p=.889$) nor for Betty (standardized $\beta=.003$, $t(37)=.023$, $p=.982$) contributed significantly to their respective models. While an understanding of human intentional thought is related to effective use of Betty's Brain, there appeared to be no link between beliefs about Betty's or the computer's mental representations and the ability to learn from the teachable agent.

Overall, students gave very similar predictions for Betty and the computer. However, there was a nonsignificant trend for students who interacted with Betty to give more dissimilar ratings of intentionality to humans and machines, implying that interacting with Betty highlighted the differences between human and mechanical agents. The person-machine difference (the intentionality assigned to the person minus the mean of the intentionality assigned to the computer and to Betty) was larger in the experimental condition (26%; $t(40)=4.008$, $p<.001$) than in the control condition (14%; $t(31)=2.103$, $p=.044$), although this contrast was not significant ($t(71)=-1.215$, $p=.228$).

4 Conclusion

We found that 7th grade students did not make more human-like intentional predictions for a teachable agent than they did for a computer. These findings mirror patterns of intentional attributions previously found in adults, but differ from adults who were asked to monitor the mental states (i.e., preferences) of a mechanical agent [6]. One possibility is that the human-agent interaction involved in using the Betty's Brain system did not sufficiently encourage students to actively monitor Betty's mental states. The students may not have considered her to be human-like (an assertion reinforced by the behavioral prediction results) but rather considered her to be a simple tool for use in completing their lessons. Many students in this study may have either required more prompting to consider Betty's mental state or alternative tasks that would have encouraged them to consider Betty's thought processes more deeply.

An alternative explanation is that children and adults are differentially influenced by mental state monitoring when attributing intentionality to a non-human agent. The students may have sufficiently monitored Betty's mental states but did not go on to infer that Betty's behavioral motivations should be any different than a computer's. By this explanation, while children and adults may both be capable of considering a nonhuman agent to have human-like mental states, adults are more likely to generalize these beliefs to other tasks (i.e., questionnaire responses).

Additionally, we found that a general understanding of agency did predict success in learning domain content in the Betty's Brain system, perhaps because students with a more elaborate theory of mind were more likely and better able to monitor the agent's knowledge states. However, this relationship only holds when considering the behavioral predictions made for the human, not the behavioral predictions made for Betty or the computer. It is possible that only the human predictions are a good general measure of the kind of social skills, or intelligence, that theory of mind depends on. The students have had more practice attributing mental states to humans, and this skill may have made them more sensitive to the metacognitive activities promoted by the Betty agent, even if students did not directly attribute intentional representations to the agent.

In this study, we found not only that children's and adults' beliefs about agent intentionality are similar, but that these beliefs can be used to predict success when learning from an artificial teachable agent. This suggests that training about agency and even ideas about intentionality could help children use teachable agent systems. Of course, it is possible that the children who gave highly intentional predictions for humans did so because of a relatively broad social intelligence that would not be strongly modified by any specific limited-duration experience. On this view, training might not be particularly effective. Instead, it might be helpful to modify the teachable agents to provide more, or fewer, cues about intentionality, depending on children's basic social-cognitive skills. In either case, we believe that understanding basic concepts about agency will be an important part of the successful design and implementation of interactive agent-based learning environments.

Acknowledgments. This material is based upon work supported by the National Science Foundation under grants #0826701, #0904387, and #0633856.

References

1. Biswas, G., Leelawong, K., Schwartz, D., Vye, N.: Learning by Teaching: A New Agent Paradigm for Educational Software. Appl. Artif. Intell. 19(3), 363–392 (2005)
2. Gopnik, A., Wellman, H.M.: Why the Child's Theory of Mind Really is a Theory. Mind Lang. 7(1-2), 145–171 (1992)
3. Dennett, D.C.: Consciousness Explained. Little Brown & Co., Boston (1991)
4. Searle, J.: Minds, Brains and Science. Harvard University Press, Cambridge (1986)

5. Woodward, A.L.: Infants Selectively Encode the Goal Object of an Actor's Reach. Cognition 61(1), 1–34 (1998)
6. Levin, D.T., Saylor, M.M., Killingsworth, S.S., Gordon, S.M., Kawamura, K.: Tests of Concepts About Different Kinds of Minds: Predictions About the Behavior of Computers. Robots, and People (under Review)
7. Levin, D.T., Killingsworth, S.S., Saylor, M.M.: Concepts About the Capabilities of Computers and Robots: A Test of the Scope of Adults' Theory of Mind. In: HRI 2008 Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, pp. 57–64. ACM, New York (2008)
8. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers. J. Soc. Issues. 56, 81–103 (2000)
9. Kuhlmeier, V.A., Bloom, P., Wynn, K.: Do 5-month-old Infants See Humans as Material Objects? Cognition 94(1), 95–103 (2004)
10. Spelke, E.S., Phillips, A., Woodward, A.L.: Infants' Knowledge of Object Motion and Human Action. In: Sperber, D., Premack, D., Premack, A. (eds.) Causal Cognition, pp. 44–76. Clarendon Press, Oxford (1995)
11. Bloom, P.: Intentionality and word learning. Trends Cogn. Sci. 1(1), 9–12 (1997)

A Comparison of Children's and Adults' Retrieval Performances and Affective Reactions When Using a Conventional Interface and an Information Visualization Interface

Andrew Large and Jamshid Beheshti

School of Information Studies, McGill University, 3661 Peel
Montreal, Quebec H3A 1X1, Canada
{andrew.large,jamshid.beheshti}@mcgill.ca

Abstract. Reports on an exploratory study of the performance and affective responses of children compared with adults when using a conventional tree-structured interface and an information visualization interface to identify subject terms embedded in a hierarchical subject taxonomy. It is part of a larger project to investigate the efficacy of information visualization as a means to enhance web-based subject taxonomies. The study found that adults were more successful at locating terms than children, but that there was no significant difference in the times taken to complete a successful search by children and adults. It also found that the affective reactions to the conventional interface versus the information visualization interface were very similar among the children and the adults.

1 Introduction

This paper presents an exploratory comparison of children and adults when faced with an identical retrieval task: to locate four terms from a hierarchically structured taxonomy that has been embedded into two different interface designs. The first interface adopts a conventional approach by presenting at each hierarchical level of the taxonomy an alphabetical list of terms that are the children of the parent term. The second interface employs an information visualization based on a radial design to present exactly the same taxonomy. The objective of the study, part of a wider investigation into information visualization, was to find out whether the children performed in a similar or a different way from the adults when undertaking the task and whether their affective responses likewise were comparable or otherwise. Relatively few studies have tried to compare adult with children's behavior when faced with a retrieval task, a notable exception being that of Bilal and Kirby in their study of web-based information seeking [1]. Slone [2] did observe and interview 11 children or teenagers of various ages and 20 adults who were users of a web-based online catalog in a public library, but did not sharply identify the responses from the children, teens and adults respectively, although reporting that motivation and experience influenced specific information-seeking behavior.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 590–598, 2011.

2 The Interfaces

Two interfaces were designed for this study. The conventional interface, "Lists", displayed on the first screen terms that are at the top level of the structure (see Figure 1); clicking on any term took the user to the screen displaying all terms at the next (second) level of the taxonomy related to this term, and so on (see Figure 2).

Fig. 1. List Interface, Top Level

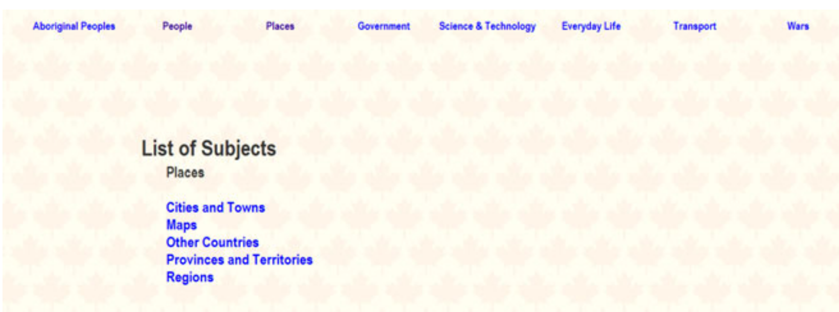


Fig. 2. Second level of "Places" hierarchy in List

The information visualization, called "Circles", was based on a 2-D radial layout technique [3] in which the position of a node relative to the center of the circle is based on the depth within the underlying tree structure. Terms at the first level of the taxonomy were displayed as a series of circles (see Figure 3). The second level was accessed by clicking on one of the circles, thereby revealing a new set of terms, grouped within the circle's circumference (see Figure 4). Terms at the third and fourth levels of the taxonomy were successively revealed in excentric clusters [4] as the user swept the cursor around the circumference (see Figure 5): as the cursor continued its sweep, earlier terms disappeared as new ones were revealed.

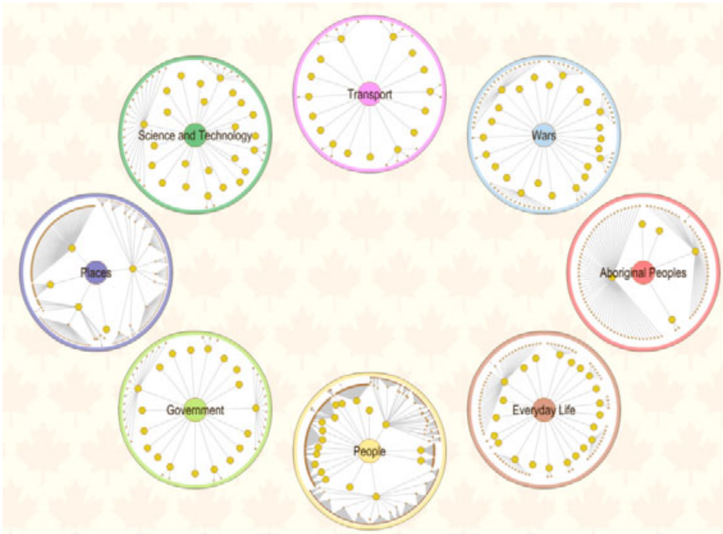


Fig. 3. Starting point of visualization navigation in Circle: the eight main topics

The information visualization interface emerged from a low-tech prototype designed over a number of sessions by an intergenerational design team of adults and children [5-6].

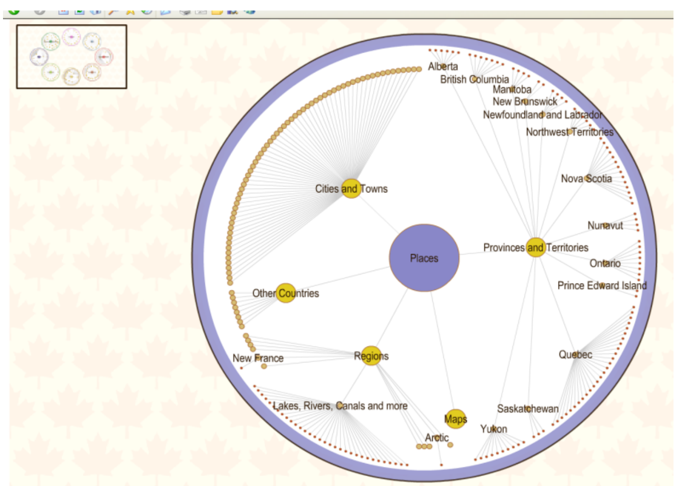


Fig. 4. Radial layout of the “Places” branch of Circles

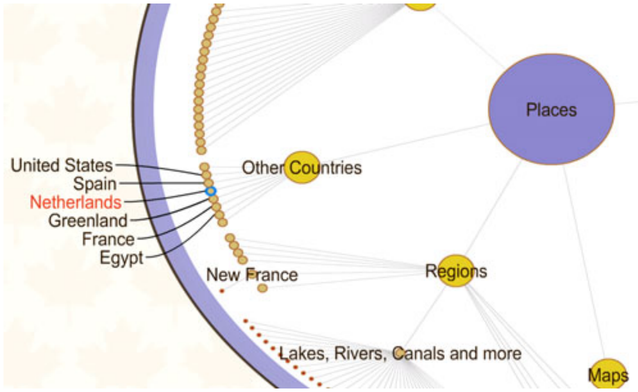


Fig. 5. Excentric Labeling (the cursor is selecting "Netherlands")

3 Methodology

3.1 The Students and Task

The participants were 13 child volunteers (7 boys and 6 girls) aged 10 to 12 years from an elementary school and 17 volunteer graduate students. They were asked individually to locate in both Lists and in Circles four individual terms, two at the third level and two at the fourth level of the taxonomy which in total comprised 1397 terms. The four individual terms – Canoes, Vaccines, Eastern Townships and Mike Myers - were chosen for this test because they were likely to be familiar to any participant, child or adult, undertaking the test. Each term was printed on a plain card, along with the following brief description:

- Search 1: Canoes – a way to travel by water
- Search 2: Vaccines – also called "shots", a way to prevent you from catching diseases
- Search 3: Eastern Townships – about one hour's drive from Montreal
- Search 4: Mike Myers – a Canadian film star

The terms were located in the hierarchical structure of the taxonomy as follows:

- Transport – Boats – *Canoes*
- Science and Technology – Medicine – *Vaccines* (or alternatively Science and Technology – Inventions – *Vaccines)*
- Places – Provinces and Territories – Quebec – *Eastern Townships*
- People – Jobs and Occupations – Actors – *Mike Myers*

3.2 The Procedure

The procedure was pre-tested with five graduate students, as a result of which minor adjustments were made to the pre- and post- questionnaires as well as the observation checklist. Detailed instructions for the research assistants were prepared and the research assistants were given a short training session to ensure that the procedure would be executed smoothly and consistently in the evaluation.

The purpose of the evaluation was briefly explained to the participant by the research assistant, following a script to ensure consistency. It was emphasized that the interfaces and not the participants were being evaluated. The students then completed a short pre-test questionnaire to collect biographical information.

This was followed by the evaluation itself. Each student was asked to locate, one by one, all four terms in the taxonomy, two in the List interface and two in Circle. The research assistant noted the student's affective behaviour as well as any obvious problems, comments or questions raised by the student on a checklist. The four terms were rotated to ensure that they were searched by the same number of students on List and Circle, and that on each interface they were searched equally as the first and as the second term. To avoid student frustration, it was decided that the duration of any search would be limited to a maximum of five minutes, after which the student would be shown how to find the term and the search would be judged a failure. Furthermore, research assistants were instructed to propose search termination to any student who appeared to be very frustrated with the task (introduced especially with the young students in mind). Successful location of a term was confirmed by an onscreen message. The on-screen activity of the students as well as any spoken commentary by either student or research assistant was captured using Camtasia software with the audio feature enabled.

Finally, each student answered six questions on a post-questionnaire to gather affective reactions to the two interfaces:

- How did you like the way words were listed?
- How did you like the way words were shown in Circles?
- Which one was easier to use?
- Which one was more fun to use?
- Which was faster to use?
- Which would you prefer to use to find a word?

3.3 Data Analysis

Two retrieval measures were applied: success or failure in locating a term correctly in the taxonomy, and in case of success, the time taken to accomplish this. Affective responses to the interfaces were measured by participants' responses to the four questions on the post-questionnaire using a five-point Likert Scale.

The 30 elementary school and graduate students each completed four searches, for a total of 120 searches; 60 were undertaken on the List interface and 60 on the Circle interface. In the case of one search, the research assistant failed to activate Camtasia; the data analysis therefore was applied to 51 searches by children and 68 searches by adults for a total of 119 searches. Camtasia searches were analyzed to determine whether the student successfully completed the search task and if so, the length of time taken to locate the sought term. The start time was noted by a research assistant when the student clicked on the opening menu screen to select either List or Circle. The end time was noted as the second when a screen was displayed saying "Bingo! You found the word". The responses to the six questions on the post-questionnaire as well as the research assistants' written observations were analyzed to identify the students' affective reactions to the two interfaces.

4 Results

4.1 Search Success or Failure

The major retrieval difference between children and adults was in failure/abandonment rates. Of the 51 term searches undertaken by the children, 11 (21%) were unsuccessful; none of the 68 adult searches were unsuccessful. In nine instances (82%) these failures were encountered in trying to find one or other of the level-four terms in the taxonomy rather than a level-three term. In 8 instances the child could not find the term in the 5-minute time period allotted; in the remaining 3 cases the research assistant proposed termination of the search because the child appeared very frustrated. There was no significant difference between the two interfaces as regards these failed and/or uncompleted tasks.

4.2 Search Completion Times

The situation was different when comparing all adult search completion times with those registered by the children who did successfully complete a search. As Table 1 shows, the descriptive statistics do not reveal major differences between children and adults. A two-sample T test confirms this result (List, t=0.246, df=52, p=0.807; Circle, t=0.307, df=53, p=0.760).

Table 1. Times to complete search (seconds)

Interfac	Particip	N	Mean	Median	Std. Dev.	Min.	Max.
List	Children	20	67.5	49	52.11	12	190
	Adults	34	63.9	50	52.84	10	258
Circle	Children	21	98.3	69	76.44	17	279
	Adults	34	92.4	73.5	64.10	13	329

List Interface. There is no significant difference between the mean times of adults and children when completing the searches on the List interface (t=0.246, df=52, p=0.807). The Mann-Whitney U Test confirms this result (p=0.768).

Further analysis by search shows no significant differences between children and adults in the times they spent completing the searches (Table 2). Due to small sample sizes, we are reporting the median times and we utilized the Kruskal-Wallis nonparametric tests here.

Table 2. Differences and similarities of search times between children and adults

Task	Participants	N	Mean	Std. Dev.	Median	p
Canoe	Children	7	65.6	57.04	39	0.271
	Adults	8	29.0	15.94	21.5	
Vaccine	Children	4	24.3	7.72	22.5	0.315
	Adults	9	36.6	23.21	30	
Eastern Township	Children	4	122.5	58.01	119.5	0.355
	Adults	9	100.4	68.09	75	
Myers	Children	5	60.8	22.16	52	0.242
	Adults	8	88.3	47.40	83	

Circle Interface. There is no significant difference between the mean times of adults and children when completing the searches on the Circle interface (t=307, df=53, p=0.760). The Mann-Whitney U Test confirms this result (p=0.815).

Further analysis by search shows no significant differences between children and adults in the times they spent completing the searches (Table 3). Due to small sample sizes, we are reporting the median times and we utilized the Kruskal-Wallis nonparametric tests.

Table 3. The differences and similarities of search times between children and adults

Task	Participants	N	Mean	Std. Dev.	Median	p
Canoe	Children	6	101.2	90.35	74	0.906
	Adults	9	91	57.52	75	
Vaccine	Children	7	110.2	106.21	57	0.487
	Adults	8	56.5	27.33	65	
Eastern Township	Children	4	75.3	18.84	76	0.865
	Adults	8	109.6	97.46	70	
Myers	Children	4	96	36.94	87.5	0.643
	Adults	9	110.4	52.17	94	

Noteworthy are two significant differences, one between children and adult time variances on the List interface when performing the Canoe search (Levene's Test for Equality of Variances, F= 9.420, p=0.009), and the other between children and adult time variances on the Circle interface when performing the Vaccine search (Levene's Test for Equality of Variances, F= 12.194, p=0.004). These results show that in both cases children were significantly more heterogeneous than adults in performing the tasks.

In terms of affective reactions, 71 percent of adults and 62 percent of children preferred Lists, whilst 59 percent of adults and 62 percent of children agreed that Circles was more fun to use.

5 Discussion and Conclusions

The specific task imposed in this study was deemed suitable because it seemed to be one that could equally be attempted by children and adults without a bias towards one group or another. Likewise the two interfaces, though based on ideas generated by intergenerational teams including children, were not specifically intended to be "child-appropriate" (in the Bilal and Kirby study [1] the search interface had been designed specifically for children).

The elementary school students (children) and graduate students (adults) differed significantly in one major respect: their ability to successfully complete their tern searches within the allotted time period (5 minutes). All the adult searches were successful, whereas 21% of the child searches were not completed. It could be that the time limit was too short, but 5 minutes seemed a lengthy time to find a term from within a taxonomy with a breadth of 8 categories and a depth of four levels. At any rate, it is impossible to speculate as to whether any of the failed child searches might have been completed with even more time available. And outside of the experimental conditions in which this study was conducted it seems unlikely that any user would have devoted more than 5 minutes to locating a term in a taxonomy as a prelude finding information within an actual database.

Excluding failed searches, there was no significant difference in the mean search time for all participants for each interface: Lists ($t=0.246$, $df=52$, $p=0.807$); Circles ($t=0.307$, $df=53$, $p=0.760$). However, children's search times were significantly more heterogeneous than adults when using Lists ($F=9.420$, $p=0.009$) and Circles ($F=12.194$, $p=0.004$) for the two level-three tasks. For all participants, Lists performed significantly better for one of the level-three searches compared with one of the level-four searches, whereas with Circles no such significance was found between searches.

The affective reactions to the two interfaces were very similar between the children and the adults both in terms of which interface they preferred - with a majority opting for Lists in both cases – and which they thought most fun to use – this time with majorities in support of Circles.

Caution must be exercised in generalizing from one small-scale study using one conventional and one visualization interface. Nevertheless, when undertaking on these two interfaces four searches that can be considered appropriate in terms of difficulty for both children and adults, these two user groups differed in their ability to complete successfully the searches, This finding is in line with that of Bilal and Kirby [1], in their study of the similarities and differences between children (in their case from grade seven) and adults (graduate students in information science) as they looked for information on the Web using a children's portal. They found that the adults were more successful than the children in finding answers to a factual question although they identified many similarities in the behavior of these two very different age groups. Likewise, in our study when comparing the adults with just those children who were successful in their searchers, no significant differences were found between adults and children in search times.

It must be emphasized that the task itself involved locating terms within a subject taxonomy, and the results cannot be indiscriminately extended to an information-seeking task such as finding information on the Web. It would be interesting to explore further this latter point, though the problem is to identify search questions that might reasonably be tackled by children and adults alike without giving one group an inherent advantage over the other.

Acknowledgements. The "Lists" interface was programmed by Marni Tam, and the "Circles" interface by Ian Clement; their expertise is gratefully acknowledged, as is the help of our research assistants. The research could not have been conducted without our children and adult volunteers. Finally, the Social Sciences and Humanities Research Council (Canada) is thanked for its funding of this project.

References

1. Bilal, D., Kirby, J.: Differences and Similarities in Information Seeking: Children and Adults as Web Users. Information Processing & Management 38, 649–670 (2001)
2. Slone, D.: The Influence of Mental Models and Goals on Search Patterns during Web Interaction. Journal of the American Society for Information Science and Technology 53, 1152–1169 (2002)
3. Eades, P.: Drawing Free Trees. Bulletin of the Institute of Combinaiorics and its Applications 5, 10–36 (1992)
4. Fekete, J.D., Plaisant, C.: Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh, Pennsylvania, pp. 512–519 (1999)
5. Large, A., Nesset, V., Tabatabaei, N., Beheshti, J.: Bonded Design Re-visited: Involving Children in Information Visualization Design. Canadian Journal of Information and Library Science 32, 107–139 (2008)
6. Large, A., Beheshti, J., Tabatabaei, N., Nesset, V.: Developing a Visual Taxonomy: Children's Views on Aesthetics. Journal of the American Society for Information Science and Technology 60, 1808–1822 (2009)

Following the Signs: Children's Use of Visual Cues to Facilitate Website Evaluation

Valerie Nesset

Department of Library & Information Studies, University at Buffalo (SUNY), 544 Baldy Hall, New York 14260-1020 USA
vmnesset@buffalo.edu

Abstract. The paper discusses how research investigating reading strategies can inform web usability design for children by identifying visual cues or 'signposts' that facilitate reading and subsequent evaluation of websites. Using data obtained from a broader study of information-seeking behavior of third-grade students conducted in Montreal, Quebec, Canada the paper reports on the students' use of visual cues when reading and evaluating websites in their search for relevant information. It was discovered that such cues as bolded text, the chunking of textual passages, color, and the thoughtful placement of images all were factors that facilitated quick and efficient evaluation of websites.

1 Introduction

Over the past 15 years and especially with the ever-increasing accessibility and popularity of the Internet both at home and at school, research into online information-seeking behaviors has steadily risen. Yet, even with this increase of studies, the information-seeking behavior of younger children still remains a relatively understudied area. Research has shown that children comprise a user group that is unique in its information needs and information-seeking behavior [1], [2], [3] meaning that they likely do not employ the same strategies as adults when evaluating websites. Indeed, research into children's intellectual behavior has shown that younger elementary school-aged children rely heavily on visual cues [4], [5], [6]. While previous LIS research has also demonstrated the importance young children place on visuals to help them perform such tasks as searching an OPAC [7] or locating a book on a shelf [8] there is a conspicuous lack of research that explores how visual cues could help with website evaluation.

This paper will investigate how third-grade children (aged eight and nine years) can make use of visual cues or "signposts" in order to accelerate online reading and comprehension while at the same time reducing the amount of in-depth reading. Some of the findings arise from a broader dissertation study of the information-seeking behavior of third-grade students conducted in an elementary school located in a suburb of Montreal, Quebec, Canada [9].

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 599–606, 2011.

1.2 Theoretical Frameworks

Since effective evaluation of websites is enhanced by knowledge of specific reading strategies on the part of the user (e.g. skimming), and is also impacted by the usability and/or design of the site (e.g. font style and size), it is appropriate to investigate studies in both areas. What reading strategies have been shown to be effective in reading online text? What website design features aid in improved usability for young students?

1.3 Reading Strategies

Elementary school students in the early grades (typically, grades one to three) are often termed "emergent" readers. An emergent reader can be defined as one who has some basic literacy skills but is not yet a fluent reader. Because of this lack of reading fluency, emergent readers often experience significant difficulties reading and comprehending text. While this is an issue with print materials, it can be an even larger challenge in the online environment. Usability research has shown that many people, whether children or adults, have difficulties reading large passages of online text (e.g. [10], [11], [12]). While there is a rich body of literature on reading instruction for emergent readers, there is little that focuses on online reading for this group. Fortunately, some of the studies that focus on reading instruction and the comprehension of text can be applied to the online environment.

Knafle [13] investigated several visual cues that could help young readers to detect and identify word structure. Using an experimental design in a print environment she examined 636 kindergarten to third-grade students to determine the efficacy of underlining and color as visual cues. She found that the use of color and underlining, in their enhancement of pattern similarities, did help with structure detection. By a small margin, underlining was found to be more effective than color. As a result of her findings, Knafle suggested using both color and underlining to help in reading instruction, especially with young readers. In a similar study, one group within a class of elementary school students received instruction in how to perform a "text feature walk" where they identified visual cues such as headings, bolded words, subtitles/headings, sidebars, pictures with accompanying captions, and labeled diagrams [14]. All of the students were then asked to read the text and answer review questions afterwards. The students who performed the text feature walk demonstrated a higher level of comprehension and learned more from the text compared to the students who did not receive the training. Although this was a small pilot study, it reinforces the idea that visual cues can help with reading comprehension.

Another method used in reading instruction is mental imaging. Students are encouraged to create mental images in their minds of the text they are reading. This has been shown to improve comprehension [15]. A corollary to this is picture-naming. In this method, children are asked to identify and describe an image in preparation to reading text that is related to it. However, research has indicated that with certain topics children may describe and/or name images differently than adults [16]. Thus, if the accompanying text is chosen by an adult (which would be the case in the online environment) it could be that the interpretations of the image will differ, perhaps causing problems with reading comprehension.

The rapid expansion of the Internet, the proliferation of information, and the increased efficacy of search engines in retrieving vast amounts of potentially relevant information are just some of the reasons that have highlighted the value of being able to quickly and efficiently read text. Skimming of text is a reading strategy that has been taught for many years, especially in post-secondary academic institutions where students and scholars are required to read, comprehend and evaluate large amounts of information. Yet, while recognized as beneficial, effective skimming was not considered an essential literacy skill. However, the Internet, in its offering of a plethora of information has changed the situation, making skimming a critical information literacy skill, not only for upper-level students and scholars but for others as well. Thus, instruction in how to effectively skim and comprehend text is becoming more prevalent in reading instruction in the latter grades of elementary school and onwards.

There are several strategies used in skimming text which are similar to those mentioned above. In a study investigating the practices of 11 high-level readers from the sixth grade as they searched the Internet [17], it was noted that the students made use of headings and the identification of familiar words to facilitate skimming of the text. Reading the first line of each paragraph was also a common occurrence. The latter method reinforces Pryke's [18] strategy of the "broken read" which encourages students to mentally delineate text into manageable chunks. Research into website usability has also shown the benefits of chunking text for easier reading [10], [11], [12].

2 Website Design and Usability

The increasing size and accessibility of the Internet has resulted in a proliferation of research studies examining website usability. Most of these studies are user-specific, that is, each one concentrates on a particular user group. The exception to this is the work of Jakob Nielsen who has been conducting usability studies with many different user groups for well over a decade. While much of his work concentrates on e-commerce and other adult-oriented sites, he has conducted numerous research studies with children. Recently, he has revisited some of his previous work with children [12] to see what (if anything) has changed. He found that children are still unable to differentiate between advertising features and actual content and have difficulty reading large pieces of text. He recommends use of bolded font, variations in colors or types of fonts, meaningful sub-headings, bulleted lists, and one idea per paragraph to facilitate skimming [11], [12].

According to a government report that pulls together research on website usability in order to provide a comprehensive list of design guidelines [10], important text on the screen should be emphasized. Furthermore, for faster reading, the report strongly encourages the use of black font on a white screen. The report cites evidence that the greater the contrast between text and background, the easier and quicker it is for people to read. The report provides several design guidelines to help address the readability issue. Some of these recommendations include: bolded font to delineate important concepts, bulleted lists, prioritizing content (i.e. important information first), and the use of color.

Color can be used to differentiate texts and to enhance website aesthetics [10], [12]. One innovative study [19] took this notion even farther; the study sought to discover if color was a factor in the establishment of trust and user satisfaction (both of which have been shown to influence e-loyalty) among three different cultures (Canadian, German, and Japanese). While the study was conducted with adults (90 people in their mid-twenties – 30 per country) using an experimental design, its results are relevant. The study found that color did have an impact on levels of trust and user satisfaction and furthermore, that certain colors were better than others, depending on the culture. These findings indicate that color can play a significant role in the evaluation of websites by influencing affective behavior.

Scrolling was a usability concern for all users in the mid-1990s, and especially for children who at that time had little to no computer experience. Nielsen's early research indicated that users were often unwilling to scroll "below the fold" (Nielsen defines the 'fold' as the webpage's initial viewing area) to view more content [20]. Over the past decade, however, access to computers and the Internet has increased dramatically. As a result, scrolling has become a common navigational technique with which users of all ages are more familiar and comfortable. Nielsen [21] cautions, however, that even though users are more comfortable with scrolling, it is still good practice to place the most important information in the initial viewing area.

3 The Study

The broader research study, conducted in winter 2006, was a phenomenological case study in a naturalistic setting—the third-grade classroom. The study explored the information-seeking behavior of two classes (52 students in total) of third-grade students in an English-language (French Immersion) suburban public school in Montreal as they looked for and used information for a class project on how Canadian animals survive in winter [9]. Of the 52 students, 12 (six in each of the two classes) were studied in-depth. Of the six data collection techniques that were used with the 12 sample-group students participating in the study, two have informed this paper: semi-structured interviews (transcribed) and participant observation (field notes in addition to screen and voice captures of 12 Internet search sessions (done in pairs during the allotted class computer time of a maximum of 45 minutes per week) through the use of Camtasia™ software). These data collection methods are similar to those used by Coiro and Dobler [17] in their investigation of the practices of advanced sixth-grade readers when searching the Internet, except they had the children perform a talk-aloud protocol while they searched. The advantage to using a screen and voice-capturing software, however, is that the participants do not have to painstakingly describe what they are doing, which can often slow them down and cause them to lose their focus as they strive to explain all of their moves instead of concentrating fully on the task at hand. The software records all of the participants' screen maneuvers along with their conversations, allowing them to focus on the search; in fact, after the first few minutes the children forgot that they were being recorded as they discussed with their partners their various search strategies and opinions on the information they retrieved. This allowed for a very natural and realistic process resulting in a rich data set. Analysis of the data followed a Grounded Theory approach using the Constant Comparison method which involves extensive coding of the data [22]. Coding was performed using NVIVO™ software.

4 Findings

The 12 study participants (five boys and seven girls) could all be described as emergent readers in English, mainly due to the fact that they were receiving instruction in two languages, English and French. They were unwilling to read for long periods of time on the Web and often jumped from website to website in their search for "good information". They tended to get distracted easily and did not have a long attention-span. Since the amount of time they had to search the Web was very short (usually only about 30 minutes) and their reading abilities limited, they were unable (or unwilling) to spend long periods of time carefully scrutinizing and reading large passages of text to find a few relevant pieces of information. Thus, they relied on the presence of visual signposts in order to accelerate the evaluation process while at the same time reducing the amount of in-depth reading.

One of the signposts that students found helpful was the use of bolded font to visually delineate relevant terms. Most often these terms appeared within section headings on a website and the children were able to quickly and easily match the words with their search terms. As Mary (all the names are pseudonyms) stated, "Well, I go to the website and read…like, if I'm looking for the habitat I look under the habitat and see which habitat describes most of it." Gertie was more precise in her articulation of her strategy; she relied first on the title of the website and if it looked promising, she would use other signposts, "…maybe its title is pretty good. Like, the title, I look at the title first, then I click on that and if I see, like, a lot of good things like habitat, where it lives, what it looks like, pictures, I'll probably stay on that website for quite awhile and pick out a lot of information and I'll go on another website." These two quotes reinforce the idea that while this matching process was somewhat useful, the students still had to read at least some of the body of text within the section to determine whether or not the information was useful. As for focusing her attention on key concept words that were bolded or in different font within the body of the text, Gertie was clear that she did not find this a useful strategy since she found many of the words "hard" and difficult to understand.

The students all used the presence of images as an indicator of a "good" site, likely because the images tended to be far more eye-catching than text. Often, they were disappointed. As Amy reflected, "Cause first like, I saw good pictures, I thought it would be like good ones so I went on the first page and it wasn't very good so I went on the second page and I went down and if I saw like, it looked good then I'd click on it…if it wasn't good then I wouldn't." Upon analysis of the Camtasia™ data, it is clear that while the students frequently retrieved and examined images of the particular animal they were investigating, often the subsequent textual description was not relevant to their needs. This was usually because the site being looked at was commercial (one girl spent several minutes thoroughly but unsuccessfully investigating a pet-food site for information on the Arctic fox). It should be noted, however, that the most efficient evaluations happened when an image was located close to brief passages of relevant textual information. Conversely, if there were no images accompanying the text, or if the accompanying text was not chunked in some way, the students often quickly left the page. Sometimes, even though an image was unrelated to the child's animal the student would become distracted and spend precious time looking at it. This behavior happened more often with the boys than the girls. An example of the lure of images was Bill who, when searching for information on the cougar, became distracted by a site full of images of the car with the same name. Bill and his partner spent most of the search session examining the car site rather than moving on to find information about the animal.

Text that was divided into chunks provided a broad signpost that helped the children to more easily detect relevant information. If the chunks were preceded by headings that included search terms and/or placed next to relevant images they were scrutinized even more carefully. Conversely, if a site contained long uninterrupted passages of text, even if accompanied by an image, the students would often abruptly leave it before they had read a word.

In the interviews and in the search sessions, although the children did not explicitly mention color it did appear to have at least some influence on their information-seeking behavior. Review of the search sessions revealed that different colored text was considered in the same way as bolded font; that is, the children seemed to perceive it to be of greater importance and were drawn to it first. The same can be said about the use of different fonts to delineate concepts—although the children did not explicitly talk about them, they did seem to notice them. Sometimes, however, the font style was difficult for the children to read and hindered rather then helped them.

In terms of scrolling, the children did not seem to mind scrolling up and down pages although they sometimes ran into mechanical difficulties when there was more than one scrollbar on the screen. Despite these difficulties, they quite happily scrolled through pages of search results and websites. Indeed, the downside was that they typically scrolled (using the scrollbars provided) up and down a site much too quickly, often overlooking signposts that potentially could have led them to relevant information.

5 Discussion

The findings of the study reinforce many of the results of previous research into reading and web usability. The students made use of visual signposts such as bolded font in titles and subheadings, chunked text, images, and color to facilitate and accelerate their reading and subsequent evaluation of website content. Bolded font facilitated the students' searches by delineating important concepts. This feature was even more useful when the bolded words were a direct match to the search terms used. Due to lack of empirical evidence, it cannot be said with certainty that the use of color or different fonts to delineate concepts was as effective although they did appear to draw the students' attention. Some fonts (e.g. cursive script), however, were difficult for the children to read as they had no prior experience with them.

The placement of images was an important factor that influenced the children's efficiency and speed when evaluating a site. A well-placed image would immediately draw a student's attention and if it was accompanied by manageable chunks of relevant text, the student was able to extract information quickly and easily. More typical situations, however, included sites that contained images that drew the student's attention even though they were unrelated to the intended search and/or sites that contained relevant images but were quickly passed over due to the daunting presence of long passages of text.

The chunking of text was an effective visual tool. Its most useful feature was that text presented in this manner appeared to be perceived by the children to be easier and more interesting to read than did long un-demarcated passages of text. The latter, when viewed, usually prompted the child to leave the page in search of more accessible information.

Color, although it appeared to be less of a factor than some of the other visual signposts, still wielded some influence on the students’ perceptions of a site. They were drawn to differently colored text, although they did not always equate it with importance (‘good information’) as they did with bolded font. Differently sized and shaped fonts also would draw their attention but more for their aesthetics than content.

Scrolling, although not a navigational impediment as shown in earlier research, sometimes hindered the students’ awareness of visual signposts. Often the students would scroll so quickly up and down the pages that they missed visual cues that indicated potentially valuable information. This indicates that web designers, in order to attract and keep the attention of young users, should try to include the most important content above the fold, perhaps in summary form accompanied by images.

6 Conclusions

Thanks to increased access, today’s young students are becoming more and more web-savvy in terms of basic navigation and other computer skills. This increased access to the Web, however, does not seem to translate into improved online reading skills. This study investigated 12 third-grade students to discover how they made use of visual cues to help them to more quickly and efficiently evaluate website content without having to engage in in-depth reading. It was determined that bolding font and using color to delineate important concepts, chunking text through the use of headings and white space, embedding images within shorter passages of text, and ensuring that important information is presented in the initial viewing area to discourage indiscriminate scrolling, were all effective signposts that facilitated quick reading and evaluation of potentially useful sites. Conversely, sites that contained elaborate scripts, long passages of uninterrupted text, headings that did not contain familiar search terms, and sites that required extensive scrolling were shown to be detrimental to successful reading and evaluation. Web designers, especially those who are creating websites for children, would be well advised to take these things into consideration when presenting content. The inclusion into webpages of meaningful visual cues that will encourage children to “follow the signs” will make for satisfying and successful information search sessions.

References

1. Walter, V.A.: The Information Needs of Children. Advances in Librarianship 18, 111–129 (1994)
2. Shenton, A.K.: Children’s Information Needs: Why do we Know so Little? Library and Information Update 3, 30–31 (2004)

3. Shenton, A.K., Dixon, P.: Issues Arising from Youngsters' Information-Seeking Behavior. Library and Information Science Research 26, 177–200 (2004)
4. Kail, R.V.: Children and their Development, 3rd edn. Pearson Education, Upper Saddle River (2004)
5. Piaget, J., Inhelder, B.: The Psychology of the Child. Basic Books, New York (1969)
6. Siegler, R.S.: Children's Thinking, 3rd edn. Prentice Hall, Upper Saddle River (1998)
7. Borgman, C., Hirsh, S., Walter, V., Gallagher, A.L.: Children's Searching Behavior on Browsing and Keyword Online Catalogs: The Science Library Catalog Project. Journal of the American Society for Information Science 46, 663–684 (1995)
8. Cooper, L.Z.: A Case Study of Information Seeking Behavior in 7-Year-Old Children in a Semistructured Situation. Journal of the American Society for Information Science and Technology 53, 904–922 (2002)
9. Nesset, V.: The Information-Seeking Behavior of Grade-Three Students in the Context of a Class Project. Unpublished Dissertation, McGill University, Montreal, Quebec, Canada (2009)
10. U. S. General Services Administration. Research-Based Web Design & Usability Guidelines. U.S. Dept. of Health and Human Services: U.S. General Services Administration, Washington, D.C (2006)
11. Nielsen, J.: How Users Read on the Web. useit.com-Alertbox (1997), `http://www.useit.com/alertbox/9710a.html`
12. Budiu, R., Nielsen, J.: Usability of Websites for Children: Design Guidelines for Targeting Users Aged 3 – 12 Years, 2nd edn. Nielsen Norman Group, Fremont (2010), `http://www.nngroup.com/reports/kids/`
13. Knafle, J.D.: Cues Aiding Structure Detection. Reading Research Quarterly 8, 502–523 (1973)
14. Kelley, M.J., Clausen-Grace, N.: Guiding Students Through Expository Text with Text Feature Walks. The Reading Teacher 64(2010), 191–195 (2010)
15. Olson, M.W., Gee, T.C.: Content Reading Instruction in the Primary Grades: Perceptions and Strategies. The Reading Teacher 45, 298–307 (1991)
16. Cycowicz, Y.M., Friedman, D., Rothstein, M., Snodgrass, J.G.: Picture-Naming by Young Children: Norms for Name Agreement, Familiarity, and Visual Complexity. Journal of Experimental Child Psychology 65, 171–237 (1997)
17. Coiro, J., Dobler, E.: Exploring the Online Reading Comprehension Strategies Used by Sixth-Grade Skilled Readers to Search for and Locate Information on the Internet. Reading Research Quarterly 42, 214–257 (2007)
18. Pryke, D.: Primary School Study Reading. Reading 26, 30–36 (1992)
19. Cyr, D., Head, M., Larios, H.: Colour Appeal in Website Design within and across Cultures: A Multi-Method Evaluation. International Journal of Human-Computer Studies 68, 1–21 (2010)
20. Nielsen, J.: Original Top Ten Mistakes in Web Design. useit.com-Alertbox (1996), `http://www.useit.com/alertbox/9605a.html`
21. Nielsen, J.: Scrolling and Attention. useit.com-Alertbox (2010), `http://www.useit.com/alertbox/scrolling-attention.html`
22. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publishing, Chicago (1967)

Development of Web-Based Voice Interface to Identify Child Users Based on Automatic Speech Recognition System

Ryuichi Nisimura, Shoko Miyamori, Lisa Kurihara, Hideki Kawahara, and Toshio Irino

Wakayama University, 930 Sakaedani, Wakayama-shi, Wakayama, Japan
nisimura@sys.wakayama-u.ac.jp

Abstract. We propose a method to identify child speakers, which can be adopted in Web filtering systems to protect children from the dangers of the Internet. The proposed child identification method was developed relies on an automatic speech recognition (ASR) algorithm, that uses an acoustic hidden Markov model (HMM) and a support vector machine (SVM). To extend the proposed method for use in a Web application, we used our voice-enabled Web system (the w3voice system) as a front-end interface for a prototype system. In this paper, we present an overview of the prototype system to elucidate our proposal. We also evaluate the efficacy of the proposed method in identifying child speakers by using voices captured from real Web users.

Keywords: Child speakers, Voice-enabled Web system, HMM, SVM.

1 Introduction

We propose a method for identifying child speakers, which can be adopted in Web filtering systems to protect children from the dangers of the Internet. In order to protect children from harmful Web sites including violence and sexual information, a technique for age group confirmation is necessary for various situations. While a practical age group confirmation technique using information related to human behavior, such as facial images, has been developed, our study investigates a method for automatically distinguishing between child and adult speakers using natural human utterances. The proposed method relies on an automatic speech recognition (ASR) algorithm [1], that uses a combination of an acoustic hidden Markov model (HMM) and a support vector machine (SVM).

ASR is a useful application of voice-enabled Web services. For an example, Google Voice Search enables users to make Google queries using their phones or other mobile devices. Several approaches have been investigated for speech-based age estimation on the basis of voices recorded in offline [2] or telephone-based situations [3][4]. In this study, to extend the proposed method for use in a Web application, we used our voice-enabled Web system (the w3voice system [5]) to the front-end interface of the proposed prototype system. Our Web-based system also enables us to evaluate the utterances collected from a real environment. To develop the practical

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 607–616, 2011.

Fig. 1. Prototype system

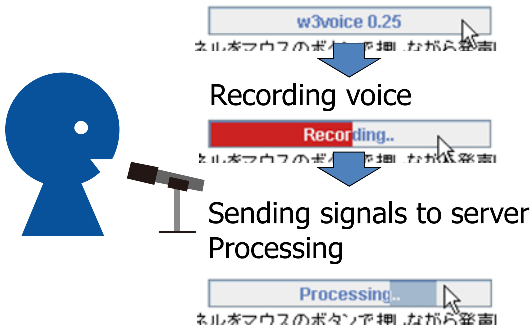


Fig. 2. Recording panel of w3voice system

speech interface, it is important to know the recording conditions prevalent in home environments. Therefore, we had to perform a large-scale collection of voices uttered at home in order to assemble research resources.

In this paper, we present an overview of the prototype system to elucidate our proposal. A two-layer algorithm for identifying children has been developed based on a combination of HMMs and SVMs. We also evaluate the efficacy of the proposed method by using voices captured from real Web users.

2 Overview of Proposed Prototype System

Figure 1 shows a screen shot of the prototype system—a Web application running on a typical Web browser. This Web page displays a normal HTML document, and a voice recording panel is provided as a component of the w3voice system [5]. As shown in Figure 2, a Web user can easily record his or her voice via the PC's microphone by clicking on the recording panel with a mouse. The recording panel also provides visual feedback as a level meter of the captured voice to the user. After recording, the captured voice signals are transmitted to our Web server where programs identify whether the speaker is an adult or a child. Finally, our system displays the result of the identification (child or adult) automatically like other cloud computing applications.

Figure 3 shows an outline of the architecture of the w3voice system. The w3voice system is freeware; it uses an open Web framework, and it can be downloaded from our Web site[1]. It can append a voice-recording interface to any Web application. Because this framework consists of a pure Java applet (client side) and CGI programs (server side), the prototype system can run on all major operating systems and Web browsers. The applet runs on the Java VM (virtual machine) of the Web browser and it records the user's voice and transmits the voice signals to the Web server. The POST method of HTTP (Hypertext Transfer Protocol) is used as the signal transmission protocol. Thus, our system can operate on broadband networks developed for Web browsing because the POST method is a standard protocol used to upload images and movie files. Web developers can easily build voice-enabled systems by utilizing our framework. The prototype system was implemented using Perl CGI scripts.

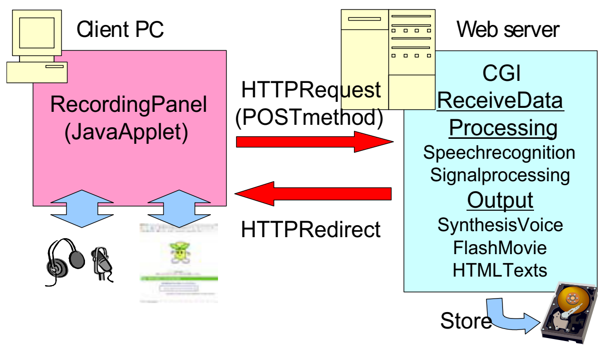


Fig. 3. Outline of w3voice system architecture

3 Proposed Method for Identifying Child Speakers Based on a Combination of HMM and SVM

To distinguish between child and adult utterances from captured acoustic signals, we have proposed a statistical model-based approach. In our previous work [1], as a simple automatic recognition method, we have tested a likelihood comparison that uses HMMs as acoustic models. A simple HMM-based classifier, which is similar to an approach used in speaker recognition, showed a relatively low accuracy of 17.8% compared to human hearing abilities in distinguishing between child and adult speakers. Our experimental results also showed that the automatic approach often made mistakes in which child samples were identified as adult female. In the age

[1] http://w3voice.jp/skeleton/

block of 15 years or more, the accuracy tended to decrease due to the influence of the cracking of voice at puberty. Because the voices of the majority of teenagers have a large variation in acoustic features, it is difficult even for a human being to identify their age groups exactly.

It is necessary to improve identification accuracies even in dealing with teenagers whose voices would change frequently. Toward this problem, in this paper, a two-layer algorithm has been developed using a combination of HMM-based ASR systems and SVMs. This technique might provide more robust identification due to handling multi-dimensional vectors of acoustic likelihood scores given by a speech recognition program, unlike a simple comparison of HMM-based method. We illustrate a procedure for the proposed method in Figure 4. The procedure of the method is explained as follows:

ASR stage (Layer 1)

(1) 2,361 short sentences uttered by 1,050 trial users of our Web site were collected using our w3voice system. To build HMM-based acoustic models, utterances were classified into 24 classes $\boldsymbol{c}$ ($\boldsymbol{c}$ = 1, 2, 3, ..., 24) according to the speakers' gender (male or female) and age group. The speakers, who were 0 to 60 years old, were divided into 12 age groups (0–5, 6–10, ..., 56–60).

(2) By using the HTK 3.4.1[2], we built 24 class HMMs ($\boldsymbol{AM_c}$, $\boldsymbol{c}$ = 1, 2, 3, ..., 24) with three-state Gaussians of 128 mixtures, using the training data in $\boldsymbol{c}$. The 12-dimensional Mel-frequency cepstral coefficients (MFCCs), delta MFCCs, and delta power were used to train the acoustic parameters.

(3) 24-dimensional vectors consisting of acoustic likelihood scores $\boldsymbol{AP_c}$ were calculated by using the speech recognition program Julius [6] with $\boldsymbol{AM_c}$ against all 2,361 utterance samples.

(4) In these procedures, as for $\boldsymbol{AP_c}$, we have used normalized values using the frame length of the captured signal.

SVM stage (Layer 2)

(5) To build a model for the SVM-based classifier, all samples prepared for training the model were divided into two classes: children and adults. For example, when the age threshold for distinguishing children from adults was 16 years, we defined the class of training samples consisting of speakers under 16 years of age as Negative (child) and the class consisting of speakers who were 16 years of age or older as Positive (adult).

(6) The 24-dimensional vectors ($\boldsymbol{AP}_1$, $\boldsymbol{AP}_2$, …, $\boldsymbol{AP}_{24}$) and the above-defined labels (Negative or Positive) were used as support vectors to train the SVM classifier.

(7) Two-value classification [Negative (child) or Positive (adult)] was performed by the TinySVM[3] with the ANOVA kernel function.

---

[2] http://htk.eng.cam.ac.uk/

[3] http://www.chasen.org/~taku/software/TinySVM/

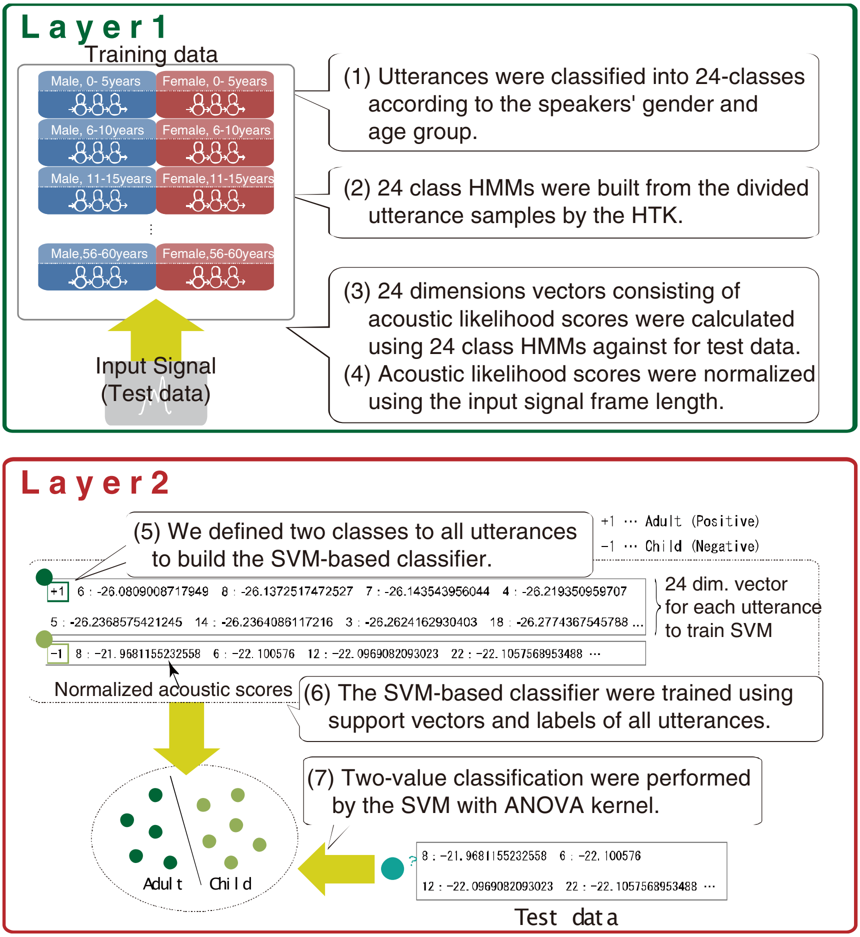


Fig. 4. Procedure of the proposed method

4 Actual Utterance Collection by Voice-Enabled Web System

Our research has obtained a large-scale Japanese voice collection including actual child voices. To collect voices in real home environments, we have prepared the voice-enabled Web site as shown in Figure 5. In order to provide an easy and friendly interface even for recording the voices of children, we developed our Web site by introducing a comic-like interface and Flash animations in the introduction section of the experiment.

The trial users, who visited our Web site using their own PCs in their homes via the Internet, uttered the answer after reading a simple question displayed on the Web pages. The questions we showed to the users were as follows:

- "Could you tell me your favorite food?"
- "Please tell me your favorite words."

Fig. 5. Screen capture of the webpage to introduce voice recording experiments

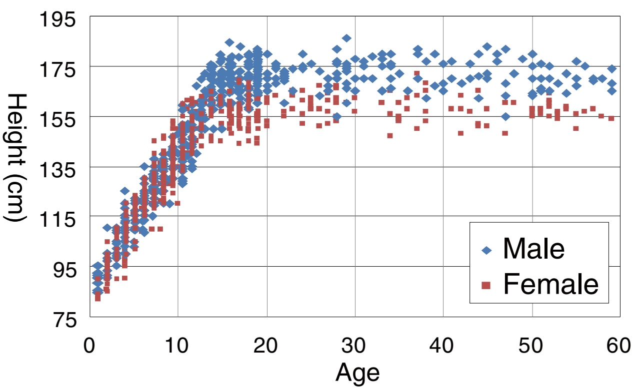


Fig. 6. Speakers' age and height in our collection

All captured voices were automatically uploaded to our Web server by the w3voice system.

In this experiment, the trial users and their parents were also requested to fill out surveys reporting their genders, ages, and hometowns. When it is difficult for child users to operate PCs and fill the reports, we requested their parents to perform vicarious operations excluding speaking acts. The trial users were recruited via the Internet monitor invitation service of the Rakuten research company in Japan. In order to cover a wide variety of attributions such as age groups and genders in the collection, the trial users were adjusted by a prior screening on the basis of a preliminary survey by the Rakuten service.

The experimental periods for collecting utterances were from February 25 to March 20 (1st operation) and December 8 to December 28 (2nd operation) 2009. As

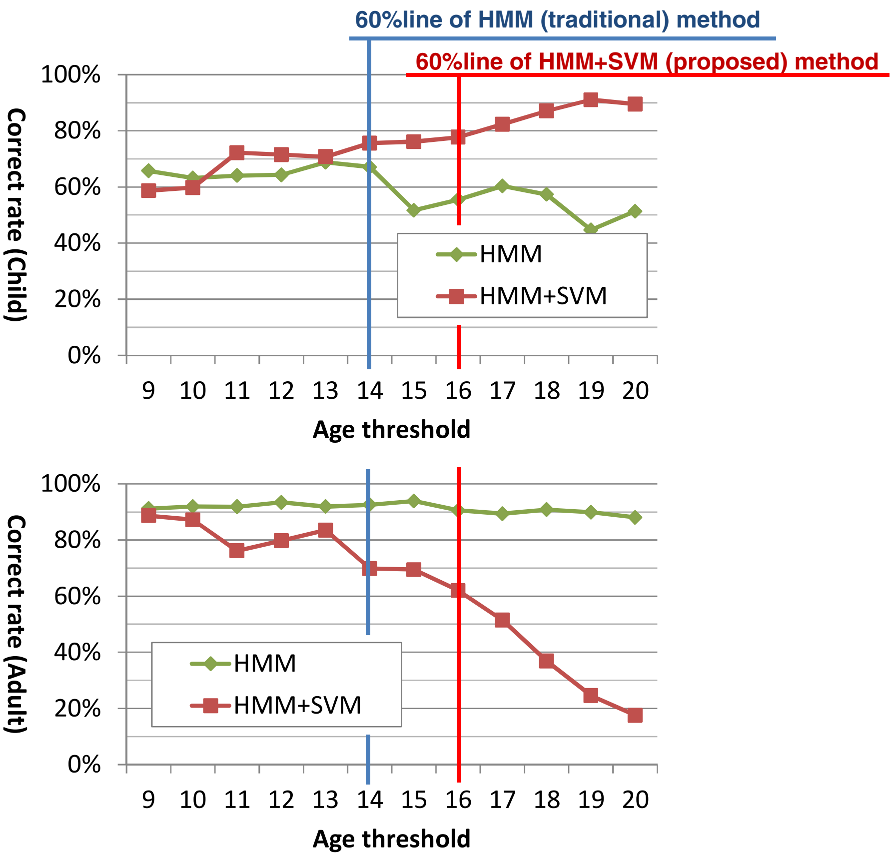


Fig. 7. Comparative results of HMM vs. HMM+SVM (Correct rate [%])

results, we have succeeded in collecting utterances from 5,778 unique IP addresses. Among them, 1,152 users had completed their voice recording and the survey on our Web site. The author confirmed the captured voices manually since these voices could include invalid recording data.

Figure 6 shows the distribution of the speaker's age and body height in our collection, which were organized based on self-reports in answer to the Web survey by the trail users. The red points indicate females, and the blue ones are males. This figure shows that we could obtain an adequate balance between the number of samples of child voices and those of the adult voices.

5 Evaluations

To evaluate the proposed child identification method, we used 2,361 utterances from our collection as the test data, and we performed a 10-fold cross-validation. The speakers used for the evaluation were excluded from the training data.

As for comparisons, we also performed tests of the traditional method using the HMM-based classifier. We have classified the collected utterances into three classes such as Child, Adult (Female), and Adult (Male), and three-class HMMs (three state, 128 Gaussian mixtures) were built. The input voice was compared to the acoustic likelihood of the HMMs and classified into the class that obtained the highest likelihood. The decoder used for identifying the class was Julius 4.1.4 [6].

5.1 Results

Figure 7 shows child and adult distinguished performances measured by the correct rate in comparisons with HMM (traditional) and HMM+SVM (proposed) method, where the horizontal (x-)axis indicates the age thresholds. We have introduced the choice of the age threshold that acts as a boundary between adults and children. The age thresholds from 9 to 20 years with intervals of one year have been examined. For example, when the age threshold was set as 15 years, the speakers in the age group of 0 to 14 years were regarded as children and speakers in the age group more than 15 years were considered as adults.

In these graphs, the (blue and red) lines of 60% correct rate are important, which could be considered according to the age threshold. The traditional HMM method could keep the 60% line of the child's correct rate between the age thresholds from 0 to 14 years. However, in the block of age thresholds of 15 years or more, the correct rate became less than 60%. In the proposed method, on the other hand, the accuracy rate of children should maintain the high success rate; however, those of adults tended to decrease. Still, the proposed method succeeded in ensuring a 60% correct rate in a block from 0 to 16 years of the age threshold. Our challenge was to get a better identification performance for young teens. Thus, the two-year increase of the 60% line has important implications. The proposed method enabled us to carry out appropriate identification of adults and children even when 16 years in the mid-teens was considered as a boundary. However, we are still faced with a low success rate of our method in overall adult identifications. It was caused by factors that samples regarded as adult utterances were reduced to be an inadequate balance of adults and children with increasing age threshold. In the future, it is necessary to develop robust algorithms to disproportionate the balance of training samples.

5.2 Comparison with the Hearing Abilities of Humans

Additionally, we also compared the hearing abilities of humans with the proposed method for distinguishing child speakers from and adult speakers. We evaluated the hearing abilities of 5 human subjects (2 males and 3 females). 260 utterances (male voice: 146; female voice: 114) were evaluated. The range of the speaker's age was from 2 to 59 years. The subjects directly listened to a recorded voice from a loudspeaker, and identified the speaker as a child or an adult.

Table 1 summarizes brief results of the evaluations (correct rate and F-measure) when the age threshold was 16 years. We achieved an accuracy of 77.8%; this implies that the proposed system can outperform the human ear in identifying the majority of teenagers, which have a large variation in acoustic features. If you would like to know detailed results for other age thresholds, please refer to Reference [1].

Table 1. Summary of HMM+SVM vs. Humans (Age threshold: 16 years old)

	Proposed method	Human hearing ability
Correct rate (%)	77.8	66.8
F-measure	0.76	0.73

6 Conclusions

In this paper, we have developed a prototype system consisting of a Web-based voice interface to identify child users based on an automatic speech recognition system. The w3voice system, which is our open framework for appending a voice input interface to a Web system, was deployed to enhance the prototype to a Web service. Now we can release the prototype system to public users via the Internet.

We have developed a new two-layer approach based on a combination of HMM and SVM in order to identify child speakers, dealing with teenagers whose voices have a large variation in their acoustic features. This paper reported the detailed performance experiments of the proposed method by using our voice collection consisting of real online users' utterances collected via the Internet. Experimental results ensured that our approach yielded improvements of performance in distinguishing adults and children even when we used the age of the mid-teens as a boundary. Thus, we found that the HMM + SVM combination method has the possibility of giving performance that exceeds human hearing abilities.

6.1 Future Works

In the future, we plan to enhance the precision of automatic identification on the basis of utilizing our voice collection. It may be possible to improve the accuracy by incorporating linguistic information in the feature vectors [7]. Further development of the prototype system is required to facilitate practical application.

Acknowledgments. This work was partially supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications, Japan.

References

1. Miyamori, S., et al.: Real world utterance collection using voice-enabled web system for child speaker identification. In: Proc. Oriental COCOSDA 2010, paper-id 46 (2010)
2. Wada, T., et al.: Investigations of Features and Estimators for Speech-Based Age Estimation. In: Proc. APSIPA ASC 2010, pp. 470–473 (2010)
3. Hempel, T.: Usability of telephone-based speech dialog systems as experienced by user groups of different age and background. In: Proc. 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems, pp. 76–78 (2006)

4. Metze, F., et al.: Comparison of four approaches to age and gender recognition for telephone applications. In: Proc. ICASSP 2007, pp. 1089–1092 (2007)
5. Nisimura, R., Miyake, J., Kawahara, H., Irino, T.: Development of Speech Input Method for Interactive VoiceWeb Systems. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 710–719. Springer, Heidelberg (2009)
6. Lee, A., et al.: Julius - An Open Source Real-Time Large Vocabulary Recognition Engine. In: Proc. Eurospeech 2001, pp. 1691–1694 (2001)
7. Nisimura, R., et al.: Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability. In: Proc. IEEE ICASSP 2004, vol. 1, pp. 433–436 (2004)

Comparison of a 3-D Expression System and a Standardized IQ Test for Children

Akihiro Suzuki, Masayuki Wajima, Takashi Kawakami, and Tetsuo Okazaki

7-15 Maeda, Teine-ku, Sapporo, Hokkaido, Japan
{r09301,wajima,kawakami,okazaki}@hit.ac.jp

Abstract. The objective of this paper is to analyze how children develop their spatial expression ability in 3-D space. We prepared a 3-D paint system that has "Translation-test", "Rotation-test" and "Paper-lay-out-test". As an experiment, we collected data by using our 3-D paint system for children between the ages of 4 and 6 in a pre-school in August 2009, January 2010, and August 2010. We conducted a survey on correlation between the system and WPPSI: WPPSI is an IQ test. At these experiments, we conducted 5 performance tests with WPPSI. The results are summarized as follows: 1) Children are able to recognize virtual 3-D space of the system, and are able to express images in virtual 3-D space; 2) The data shows the trend that spatial expression ability in space becomes better with advancing ages; 3) We obtained correlations between the system and WPPSI. The system can analyze some abilities that are difficult for WPPSI.

Keywords: Spatial vision, Spatial expression, Human-computer-interaction, Children.

1 Introduction

Various articles have been published on the process of how children in Japan develop their spatial perception. For example, these articles try to understand it by means of studying paintings[1], using toy blocks[2] and having conversations[3][4]. However, paintings depends on 2-D expression. And with toy blocks, children are highly likely to have space with different parameters. Children likely to have a different sense of gravity and children's space likely to have a different coordinate axis from our adult's space. And children's conversational ability needs a long period for development. The Wechsler Preschool and Primary Scale of Intelligence test (WPPSI)[5], one of the standardized IQ tests, has subtests to analyze spatial perception ability. But the WPPSI analyzes ability using 2-D space. Moreover, test examiners need to understand the WPPSI test, and children need to endure a test period of longer than 2 hours.

So, we feel that our approach, which uses a computer, is more effective. We prepared a 3-D paint system to measure the developmental growth of spatial expression ability in 3-D space. As far as we know, this is the first approach that studies children's spatial development using a computer. We conducted a series of experiments with 4 to 6 years old children in a pre-school, as the first step in studying

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 617–626, 2011.

the development of spatial growth. The aim of the experiment is to measure how children develop their spatial expression ability over a period of time.

2 Prepared System

2.1 3-D Paint System

We define some basic spatial expression abilities as follows:

- Operation abilities in defined 3-D space
  - ability to cognize and understand a limited 3 dimensional area
  - ability to move something to a target position in a space, or ability to draw a figure in a 3 dimensional area
- Expression abilities of relations
  - ability to cognize relations of distances, directions, and dimensions between one thing and the other
  - ability to decide distances, directions, and dimensions for the target relations

Our 3-D paint system needs functions for understanding those abilities, and needs methods of spatial expressions for understanding that children's immanent 3-D space. So, we formulated the following methods of spatial expressions in a computer, and employed those methods on our 3-D paint system. Fig.1 is a mock picture showing the stage of expression of complex perspective (4 to 7 years old) according to Higashiyama[1]. The picture has a road and a railway track from an overhead view and a house and a train from lateral view. A 3-D image can be expressed by dividing Fig.1 into various parts and to lay out the parts in a 3-D image as Fig.2. Similarly, we applied this method to the system. The system has a function that a user to lay out illustrated papers in a virtual 3-D program. We also equipped the system with other functions by means of operating cubes for understanding the operation abilities.

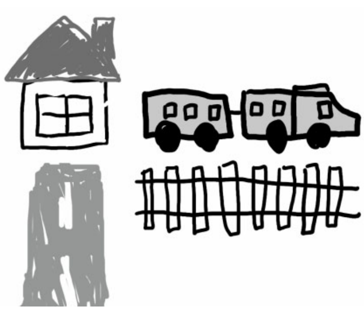


Fig. 1. A child's drawing

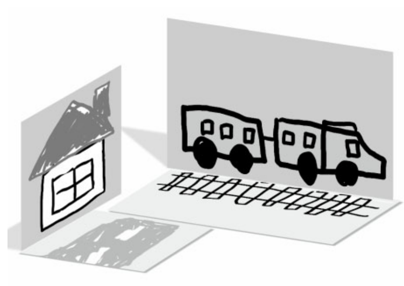


Fig. 2. Lay out in 3-D space

2.2 Overview of the System

We prepared a 3-D software paint system which meets the requirements in section 2.1. We developed the system using the C++ programming language, the system works on Windows and Mac OS X. Fig.3 shows the execution screen of the system. The function buttons are located at the top of the screen. Located at the left of the screen are a color palette, undo button and redo button, and the button that can be used to create a paper. The rest of the area is space to express 3-D images.

We applied Nintendo's Wii Remote[6] as an operation device for the system. Because, to use a mouse and a manipulator, which are commonly used in computer graphics, is hard for children who have limited or no computer experiences. The operation device for the system requires acceleration sensors and gyroscope sensors to rotate the papers, and optical sensors and buttons to translate and choose the papers. Applying Wii Remote to the system is useful in introducing the system at a pre-school or at home. Because anyone can obtain Wii Remote easily. So, we adopted Wii Remote for the system. We henceforth call Wii Remote "3-D Remote".

The system has functions of creating new papers, drawing on papers, displaying cubes for experiments, translating and rotating papers and cubes, and translating and rotating a view. And the system has functions of Translation-test, Rotation-test, Paper-lay-out-test for experiments. The Translation-test and the Rotation-test are tests for analyzing the operation abilities in 3-D space. The Paper-lay-out-test is a test for analyzing the expression abilities of relations.

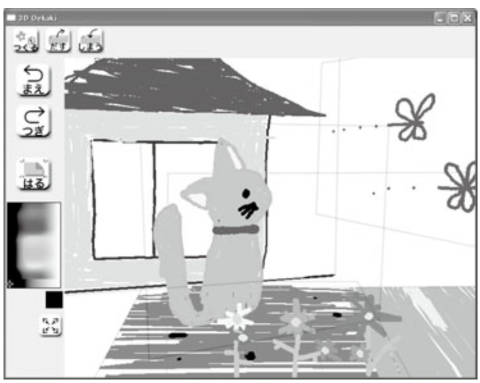


Fig. 3. 3-D paint system

2.3 Methods of Operation

Fig.4 shows a layout of the buttons of the 3-D Remote (Wii Remote), and Fig.5 shows a coordination system of the system. The system has functions of translating and rotating papers, and translating and rotating a view by using the 3-D Remote.

1. Translations: A user can translate a paper by hitting the arrow keys of the 3-D Remote. If the 3-D Remote is horizontal, a paper will translate back, forth, right or left on the X-Z plane. If the 3-D Remote is vertical, a paper will translate up, down, right or left on the X-Y plane. The threshold value is 45 degrees.

2. Rotations: A user can rotate a paper by holding the B-button and rotating the 3-D Remote. If a user rotates the 3-D Remote around the X-axis, a paper will rotate around the X-axis. If a user rotates the 3-D Remote around the Z-axis, a paper will rotate around the Y-axis.
3. Changing an object: A user can move a cursor by holding the 3-D Remote in the direction of a display and moving the 3-D Remote. And a user can change an object by pointing to an another object and hitting the A-button.

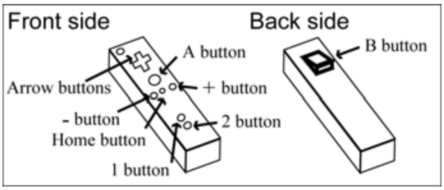


Fig. 4. 3-D Remote

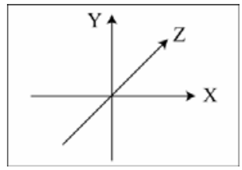


Fig. 5. Coordinate system

3 Experiment Overview

We conducted the experimental tests in August 2009, January 2010, and August 2010 in a pre-school as shown Fig.6. The examinees were 14 children 4-years-old, 17 children 5-years-old, and 11 children 6-years-old, 26 boys and 16 girls. 15 children took the experiments twice in the first experiment and also in the second experiment.

We tested Translation-test, Rotation-test, and Paper-lay-out-test for children in every experiment. In August 2010, we used the WPPSI IQ test. These experiments were conducted in individual room. We explained to them the method of operation, and tests before tests. The required time of experiments was about 20 minutes. Fig.7 shows a photo of the experiment. Fig.8 shows the process of the tests.

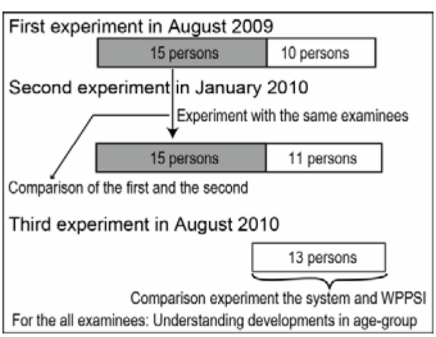


Fig. 6. Experiment plan

Fig. 7. Experiment in a kindergarten

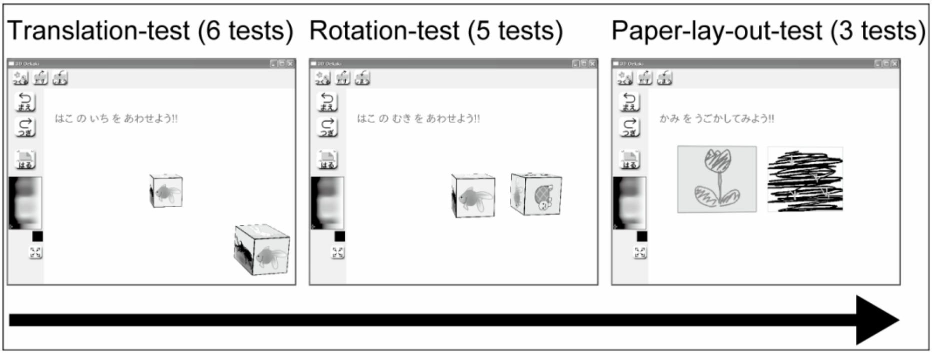


Fig. 8. Process of tests

4 Difference of Operation Ability in Space in Half a Year

4.1 Experiment Methods

We describe Translation-test and Rotation-test as follows:

- Translation-Test: Two cubes of the same size but different colors are arranged at different positions in advance. An examinee translates the highlighted cube at the center to another cube. There are 6 different tests as shown Fig.9.
- Rotation-test: Two cubes of the same size but different colors are arranged in different directions in advance. An examinee rotates the highlighted cube until the same direction is obtained as the other one. There are 5 different tests as shown in Fig.10.

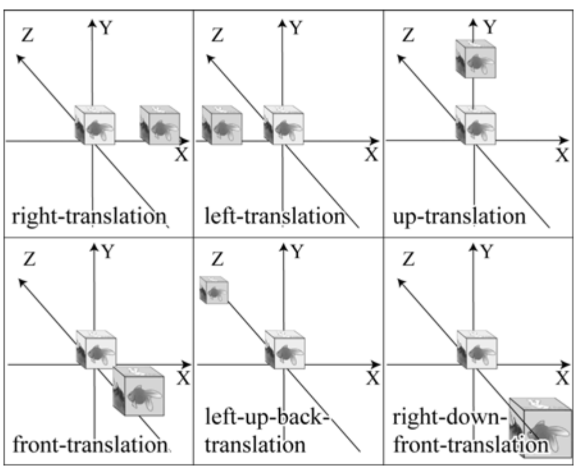


Fig. 9. Translation-test

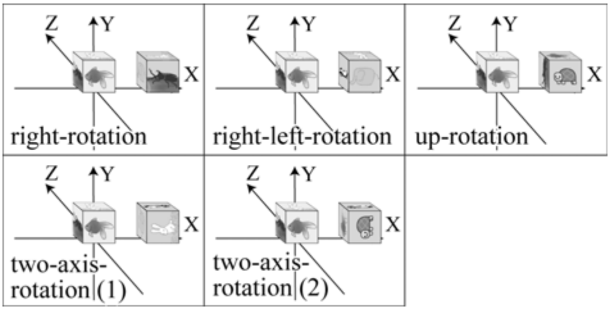
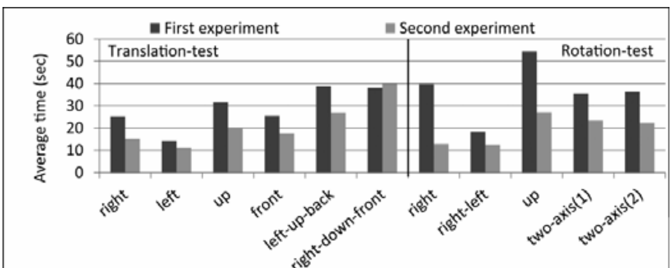


Fig. 10. Rotation-test

4.2 Experimental Result

Fig.11 shows the average times and the standard deviations of time required to complete a test, and the rates of completion. In the second experiment, most of the average times of the Translation-test and the Rotation-test are shorter than the first experiment. And in the second experiment, most rates of completion of the Translation-test and the Rotation-test are greater than the first experiment. By t-test, the average times of right-rotation, right-left-rotation, and two-axis-rotation(2) are significantly different between the first experiment and the second experiment ($p<.05$). By F-test, the average times of right-translation, up-translation, right-rotation, right-left-rotation, and up-rotation are significantly different between the first experiment and the second experiment. This result provides evidence that shows the difference of children's development of operation ability in space. Moreover, the result provides evidence that children can recognize the virtual space.



		right	left	up	front	left-up-back	right-down-front	right	right-left	up	two-axis(1)	two-axis(2)
first experiment	average time	25.3	14.2	31.5	25.5	44.5	41.3	39.7	18.3	54.3	35.6	36.3
	standard deviation	26.6	11.6	40.4	15.1	25.3	26.3	43.4	23.8	38.7	22.1	17.5
	completed rate	1.00	1.00	0.87	0.93	0.73	0.60	1.00	1.00	1.00	1.00	0.80
second experiment	average time	15.1	11.1	19.7	17.6	29.9	43.1	12.8	12.6	27.1	23.3	22.3
	standard deviation	13.5	7.5	16.5	19.0	17.1	52.9	16.2	11.7	19.3	16.4	18.0
	completed rate	1.00	1.00	1.00	0.93	0.73	0.73	1.00	0.93	1.00	0.87	1.00

Fig. 11. Average time and standard deviation in 6 month

Fig.12 shows the average times and the rates of completion of all the examinees of the Translation-test and the Rotation-test in their age-group. The average times decrease with the rates of completion increasing with advancing age. However, the average time of 5-year-old children for the Translation-test is longer than for the 4-year-old children. Because the rates of completion increase from 5-year-old, the examinees of 5-year-old that could complete the tests with longer time increased. This result shows that we can measure development of operation ability in space using the system.

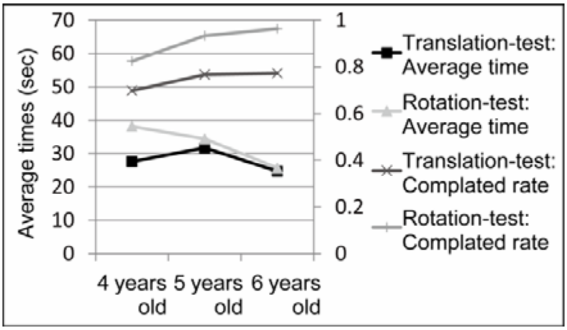


Fig. 12. Average time and rate of completion

5 Correlation of the System and WPPSI

5.1 Experiment Methods

In our experiment in August 2010, we used this system and the Japanese version of the WPSSI for the 14 children. The aim of this experiment was to validate the evidence that the system is an effective tool to measure a developmental stage of spatial expression ability.

The Japanese version of the WPPSI is an intelligence test designed for children ages 3 years 10 months to 7 years 1 months. The WPPSI provides subtests with 6 verbal tests and 5 performance tests. And it provides Verbal and Performance IQ scores, and a Full Scale IQ score. However, the WPPSI is not specialized in understanding the spatial expression ability. We conducted all 5 performance subtests of Animal House, Picture Completion, Mazes, Geometric Design, and Block Design.

5.2 Experimental Result

Fig.13 shows the examples of correlation between the required times of the system and the scores of the WPPSI. The upper graphs show correlations between average times of the Translation-test and the WPPSI, and the lower graphs show correlations between average times of the Rotation-test and the WPPSI.

In the Translation-test and the WPPSI, the average time increases with the score of WPPSI increasing. This result is the same in other all Translation-tests. By test of no correlation, Picture Completion and Geometric Design correlate significantly with the Translation-test ($p<.05$). Moreover, the Performance IQ score correlates significantly with the Translation-test. This result shows that children have an ability of systematizing a way to reach a goal in Mazes, or an ability of painting by comparing an original with their own picture in Geometric Design effect in virtual space. But, we found no correlation between the Rotation-test and the WPPSI. However, this result shows that the system can analyze abilities that are hard for the WPPSI to measure because the results of the Rotation-test improve with advancing age in section 4.2.

That is, the system obtained an equal result to the WPPSI for analyzing spatial expression ability. And the system can analyze spatial expression abilities that are hard to understand with the WPPSI.

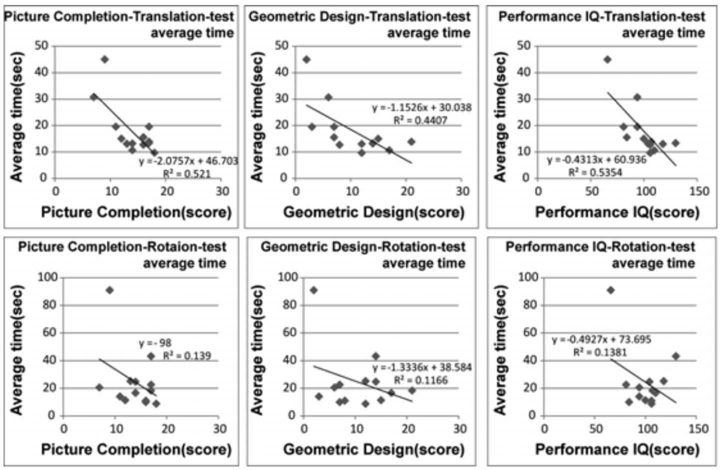


Fig. 13. Correlation between the system and WPPSI

6 Paper-Lay-Out-Test

6.1 Experiment Methods

Paper-lay-out-test is an experiment to test the ability to express the positional relation between two papers. Two illustrated papers are arranged in virtual 3-D space of the system in advance. In real space, the test examiner shows a relation to the child using the two illustrated papers. And in the virtual 3-D space of the system, the child will arrange the papers similar to the relation.

The Paper-lay-out-test consists of Parallel-lay-out-test, Vertical-lay-out-test(1), and Vertical-lay-out-test(2) as in Fig.14. We added vertical-lay-out-test(2) from January 2010.

6.2 Experimental Result

Fig.15 shows the 9 typical expression patterns that we obtained from the examinees. Table.1 shows the frequencies of expression patterns of all examinees and the frequencies of examinees of the first experiment and the second experiment, the underlined data shows the correct expression pattern.

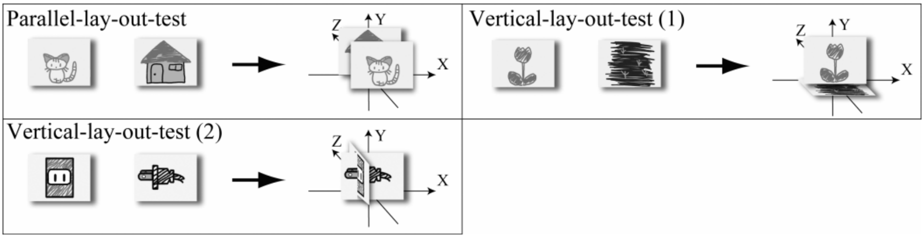


Fig. 14. Paper-lay-out-test

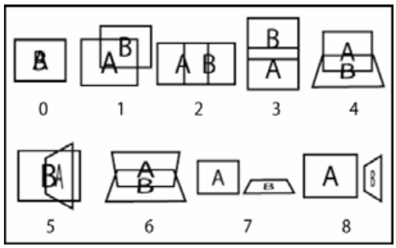


Fig. 15. Expression patterns

Table 1. Result of Paper-lay-out-test

	Parallel-lay-out			**Vertical-lay-out(1)**			**Vertical-lay-out(2)**
patterns	all	first	second	all	first	second	all
0	36	10	8	24	6	4	14
1	**8**	**2**	**1**	9	4	1	1
2	5	2	2	1	0	0	5
3	1	0	1	3	0	2	0
4	0	0	0	**14**	**3**	**5**	0
5	7	1	3	2	0	2	**16**
6	1	0	0	2	1	1	0
7	2	0	0	3	1	0	0
8	1	0	0	1	0	0	1

In the pattern 0, two papers are arranged at the same position, is highest frequency in all the tests except the Vertical-lay-out-test(2). This result is caused by a lack of understanding the experimental descriptions, a confusion of the Translation-test, or undeveloped spatial expression ability.

In the Parallel-lay-out-test, 8 examinees expressed the correct expression for the pattern 1. The rest of children expressed the pattern 2 or 5. We could not obtain significant differences between the first experiment and the second experiment. This result is caused by the following reasons: it was hard for examinees to understand whether they should arrange the papers with opening a gap or arrange the papers to the same position such as the pattern 0; or undeveloped spatial expression abilities.

In the Vertical-lay-out-test(1), 14 examinees expressed the correct expression for the pattern 4. We see that the number of the pattern 4 increases, the number of the pattern 3 and 5 as nearly correct expression increases, the number of the pattern 0 and 1 decreasing in half a year. In the Vertical-lay-out-test(2), the highest number of 16 examinees expressed the correct expression for the pattern 5.

This result shows that we are able to obtain spatial expression in 3-D space, and the system obtained data showing children's development of the expression abilities of relations with advancing age.

7 Conclusion

We prepared a 3-D paint system to understand the developmental stage of spatial expression ability, and conducted evaluation experiments using the system in a pre-school. The results are summarized as follows:

- Children are able to recognize virtual 3-D space of the system, and are able to express images in virtual 3-D space.
- The data shows the trend that spatial expression ability in space becomes better with advancing age.
- We obtained correlations between the system and WPPSI. The system can analyze some abilities that are difficult for WPPSI.

In future studies, we will continue to conduct experiments in a pre-school, and make an evaluation method to understand developmental stages of spatial expression ability from the acquired data.

References

1. Higashiyama, A.: Children's Space Recognition and It's Expression. Ann. Kobe University 83, 71–81 (1989)
2. Wechsler, D.: Wechsler Intelligence Scale for Children-III, San Antonio (1991)
3. Takeuchi, Y.: Can Children Understand Aerial Photographs as Map-like Representations? Ann. Aichi University of Education 6, 43–48 (2003)
4. Ogawa, A.: Young Children's Spatial Perspective Taking Ability: The Relationship between Children's Understanding and Production of Drawings in Two Conditions. Ann. Kyoto University 52, 412–426 (2006)
5. Wechsler, D.: Wechsler Preschool and Primary Scale of Intelligence, San Antonio (1967)
6. Ito, K., Fukuda, T.: Wii Remote. J. Transactions of the Japan Society of Mechanical Engineers 110, 908–909 (2007)

Exploring Children's Requirements for the Graphic Design of WebOPAC

Tengku Siti Meriam Tengku Wook[1] and Siti Salwah Salim[2]

[1] Faculty of Information Science and Technology,
National University of Malaysia, 43600 Bangi Selangor, Malaysia
[2] Faculty of Computer Science and Information Technology
University of Malaya, 50603 Kuala Lumpur, Malaysia
tsm@ftsm.ukm.my, salwa@um.edu.my

Abstract. This research specifically focuses on the graphic design issues involved in a WebOPAC interface for children. Children's ideas, expectations, experiences and needs were explored through their involvement in a variety of participatory design (PD) activities. The main outcomes of these activities were lists of the children's preferences and interests, sketches of design ideas and metaphors which led to the specification of the children's requirements for the graphic design of a WebOPAC interface. This research also points out the potential of information visualization (IV) and virtual reality (VR) techniques in terms of supporting the implementation of these requirements.

Keywords: Children's requirements, graphic design and participatory design method.

1 Introduction

This research focuses on the issues involved in the graphic design of a web-based information retrieval system (IRS) for children, specifically for WebOPAC. A version of WebOPAC known as Ilmu has been used by children in Malaysia to search and browse the bibliography of books in a library. Usability testing of WebOPAC has shown that poor graphic design in particular has contributed to usability problems [1]. The parameters applied in the user testing were poor graphic design, problems with navigation, poor functionality, and too much text on display. WebOPAC's graphics need to be redesigned in order to increase the usability of its existing interface, so that bibliographical information can be retrieved easily and effectively.

A detailed literature review identified three main elements of graphic design: layout, use of colors and human visual perception [2,3,4,5 and 6]. Layout is the design of an interface, which involves the screen and the window. Layout comprises an important issue in graphic design because it incorporates the use of space. It encompasses the effect and arrangement of text, objects (e.g. labels, dialog boxes, icons and menus) and the functions of the interfaces which are connected through space [7].

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 627–636, 2011.

Human visual perception is related to the issue of layout, which is related to the spacing and alignment of icons and labels in the interface. It is quite difficult to determine human visual perception during the production of a graphic design, as the preparation and placement of information change continuously. Moreover, human visual perception relies on an individual's sense of aesthetics, which differs according to their age and requirements [8]. However, human visual perception can still provide guidelines and principles for the design of a visually pleasing interface.

According to Benson [9], colors are accessories which are added to a user interface; they are not an essential aspect of the design of a user interface, especially for a web-based application, as the application will still work without the use of colors. However, the use of color is very important in graphic design, as it helps to change the user's feeling regarding or perceptions of the system interface [10].

Layout, the use of color and human visual perception are the main graphic design-related issues that need to be addressed with regard to the current WebOPAC interface. This study focuses on children's requirements as a basis for redesigning the WebOPAC interface through the use of a participatory design approach, which involves children as design partners. Section 2 of this article describes the participatory design methods employed in order to explore the children's needs. Section 3 presents the data analysis which led to the specification of requirements for the graphic design of the interface. The concluding section summarizes the results of this paper and highlights several potential techniques and technologies which could be used to support the implementation of these requirements.

2 Participatory Design Method

Figure 1 shows the participatory design approach which was used in order to explore children's requirements for the graphic design of WebOPAC. This approach involves two main activities, which are:

i. Inquiry helps children to identify their preferences regarding graphic design through exposure to existing WebOPAC systems;
ii. Sketching out ideas helps children to express their design ideas and preferences in the form of drawings.

This study involved children aged between seven and 11 years old, and they were all Internet and information technology literate.

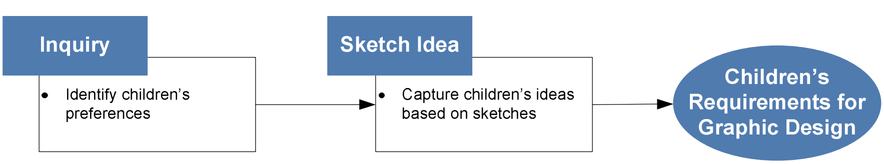


Fig. 1. Participatory design method

2.1 Inquiry

The children were introduced to two existing WebOPAC systems: Kid's Catalog Web [11] and Kids Web [12], and one digital library, ICDL [13]. They were required to retrieve bibliographical information using keywords, subjects, locations and other features provided by the systems. They then recorded their preferences and expectations on a checklist. This activity was conducted with 40 children in a computer laboratory. The checklist consisted of 15 items, focusing on the elements of layout, the use of color and human visual perception. The items were extracted from guidelines gathered using a literature review [14], [9] and [15].

Analysis and Discussion. Figure 2 shows the percentages representing the extent to which the students preferred the 15 items on the checklist. Items with percentages which were higher than 50% were selected as representing the students' preferences. As shown in Figure 2, all of the items obtained a score which was higher than 50%, signifying that the students preferred all of the items. These preferences were based on the children's experiences of using three IRSs.

Figure 3 shows the students' preferences according to the key elements of graphic design: layout, the students' visual perceptions and the use of color. It can be seen that students prefer a bright and attractive combination of colors. Students require a layout that displays information in a visual environment, such as in a three-dimensional (3D) library, leaves space between the objects (images, icons and buttons), and in which objects are displayed in a consistent and coherent way. Students require a visual perception that displays objects which are labeled in a precise and consistent way and contained animations in order to attract their attention.

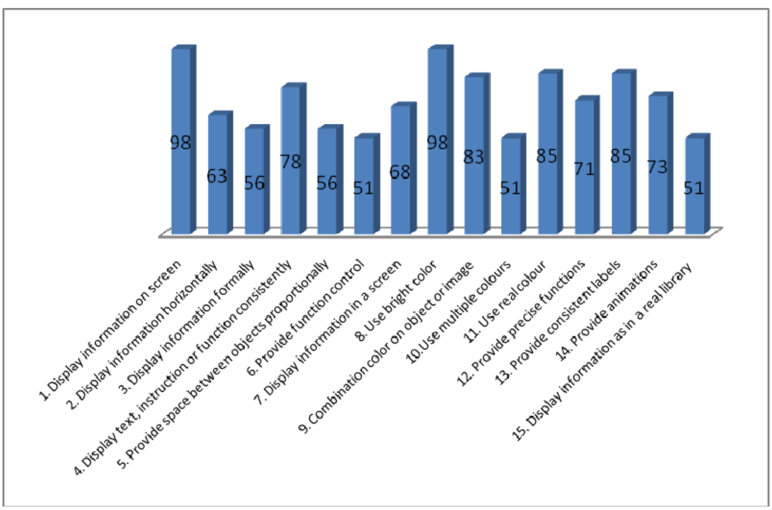


Fig. 2. Percentages of students' preferences

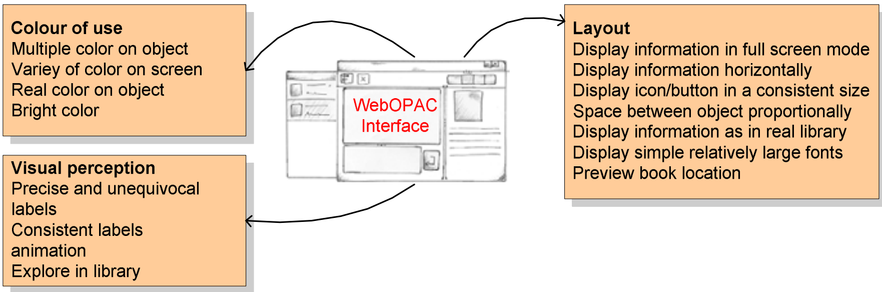


Fig. 3. List of students' preferences

2.2 Sketch Idea

The second activity was carried out with 20 children (divided into four groups). The purpose of this activity was to allow the students to generate design ideas and to display their preferences in the form of sketches. PICTIVE [16] and CARD [17] techniques were used to support these sketching activities.

Analysis and Discussion. Analysis of the four sketches showed that the children used metaphors in representing information in the form of images and graphics, as shown in Table 1. In order to gain an understanding of the metaphors produced, a walkthrough session was conducted with each group.

From the list in Table 1, one can see that there are similarities between the design ideas expressed and generated by the students, which represent their strongest preferences. The most common ideas were:

i) All of groups described the library environment and the actual space in the library which could be explored by users;
ii) Using signs as indicators of direction (except Group C);
iii) The use of symbols e.g. a book or a script showing the information content;
iv) Groups C and D drew the front entrance or resource center as a symbol that the knowledge and information in the building can be accessed by anyone;
v) Groups B and C sketched maps of Malaysia as a way of communicating the location of the books which are stored in each state;
vi) Groups B and C used superhero characters as assistants in the information search activities.

The differences between the groups were as follows:

i) Group A visualized the search for the location of books using patches of cloud, while Group C represented the search using information stored in computers;
ii) Group A used the plus and minus symbols as well as insects to indicate different subjects, while the other groups only listed the subjects in the form of words;
iii) Group D used the search box and ball games to search for bibliographical information.

Table 1. Images and graphics used in the sketches

Group	Metaphors	Information
A	Patches of cloud	Shows the location of books kept anywhere in Malaysia
	Book	Shows that a lot of information can be accessed through the interface
	Plus and minus symbols	Indicates mathematical subjects
	Insect	Indicates scientific subjects
	Signpost	Indicates signage and direction
	Tables, chairs, sofas, bookshelves and a flower pot	The library environment
B	Map of Malaysia	Showing the location of the books kept in every state in Malaysia
	Three railway carriages	Shows the loading of information in keywords, subject and location
	Car	Transporting students in order to bring children to the library
	Superhero character	Showing that assistants are essential for providing information
	Signpost	Direction and signage
	Tables, chairs, sofas, bookshelves and a flower pot	The library environment
C	Main entrance to the library	Symbolizes the building
	Map of Malaysia	Showing the location of the books kept in every state in Malaysia.
	Superhero character	Showing that assistants are essential for providing information
	Tables, chairs, sofas, bookshelves and a flower pot	The library environment
	Script	The display shows information on a sheet of paper
D	Main entrance to the library	Symbolizes the building
	Box	Indicates the process of searching for information
	Ball	Include game features
	Computers	Shows the location of and information about stored books
	Signposts	Direction and signage
	Tables, chairs, sofas, bookshelves, flower pot	The library environment

Table 2 shows the overall design ideas generated by each group in relation to the WebOPAC environment and functionalities, and the seven items extracted from the design ideas that were associated with elements of graphic design (layout, visual perception and use of color). With regard to the layout preferred by the children, this can increase the effectiveness of the children's learning because the visual environment has a well-structured space and the items are arranged in a suitable manner. In addition, the children's visual perception can be improved by the usage of labels on images as well as by the use of space, giving children the opportunity to explore whilst playing in the library. Furthermore, the colors chosen by the children for the interface were bright and varied, in order to encourage an attractive and engaging experience in a library environment.

Table 2. The children's design ideas and the items associated with the key elements of graphic design

Group	Design ideas	Items associated with graphic design
A	▪ Search by location ▪ Using the symbols on the icon/command ▪ Overview of the library ▪ Signs	▪ Search by location rather than using keywords or subjects ▪ Using the symbols on the icon/command ▪ Use of space to describe the environment in the library ▪ Animations ▪ Using signage ▪ The concept of the game ▪ Available characters
B	▪ Search by location ▪ Using the symbols on the icon/command ▪ Some animations ▪ Characters as assistants ▪ Signs ▪ Overview of the library	
C	▪ Search by location ▪ Using the symbols on the icon/command ▪ Overview of the library ▪ Characters as assistants	
D	▪ Search by location ▪ Using the symbols on the icon/command ▪ Signs ▪ Overview of the library ▪ Include game features e.g. search using game approach	

3 Requirement for Graphic Design

These lists of preferences, sketches of ideas and metaphors gained from the children (Figure 3 and Table 2) facilitated the identification of fourteen lists of requirements which served as a basis for redesigning the graphics for Ilmu's interface. Table 3 is a list of the children's requirements for the graphic design of WebOPAC.

Table 3. List of the children's requirements

CHILDREN'S REQUIREMENTS FOR THE GRAPHIC DESIGN OF THE WEBOPAC	
1. Display full information	8. Use the correct labels
2. Information arranged horizontally	9. Use labeling consistently
3. Formal appearance of the object/image	10. Provide animated functions
4. Medium amount of space between the objects	11. Use objects in a variety of colors
5. Display icons/buttons in a uniform manner	12. Use many colors on the screen
6. Display book location	13. Consistent use of the actual color of natural objects
7. Display information about the actual library	14. Use bright colors

These 14 requirements were categorized into four categories: the use of space, the organization of information, function and the use of color. These categories reflect the causes of the previous usability problems with the elements of graphic design. These categories, together with the children's requirements and examples of features that illustrate how these requirements are to be met, are depicted in Table 4.

The first category covers the spatial and the function control aspects of assisting children in searching and browsing information-based activities on the database. These aspects can also increase children's understanding of how to use WebOPAC. The second category refers to visual information which must be shown on the WebOPAC user interface in order to display objects/images/icons/buttons in a simple and economical manner. This is done in order to make the information appear uncluttered and simple. The third category covers the accessories that are used on the WebOPAC layout in order to increase comprehensibility. The fourth category involves the use of colors that are cheerful and bright in order to attract the children's attention.

The four categories and 14 requirements demonstrate the concepts of displaying information visually in a space in which children can control the functions and explore the entire space. According to Baeza-Yates [18], information visualisation (IV) and virtual reality (VR) can be applied in an overall information display that can be visualised in a space. Table 4 clearly indicates that IV and VR are capable of supporting an information display in terms of controlling functions and helping children to explore the space inside a library. Each requirement has its own feature, which can be controlled or displayed according to the function performed by the children on the items being displayed. Features that can be controlled and displayed using IV are R1, R2, R3, R4, R10 and R11. There are also features that do not require the use of a specific technique, as they can be displayed by means of the user requirement specifications.

Popular IV techniques include a zoomable user interface (ZUI), a lens [19] and tree-maps [20]. The VR technique is also effective for creating a realistic information display in a 3D environment, displayed in 360° vision and navigable using a keyboard or a mouse [10]. In addition to exploring a 3D environment, the walkthrough technique can be used in order to explore a 2D environment [21]. This technique is a combination of IV and VR, which involves a graphical information display that can support 360° visual exploration in 2D or 3D. This concept is intuitive, as it can be operated via user intuition [22].

Table 4. Requirements for graphic design

Category	Requirements	Features
(i) Use of space	R1: Display information as it is in the real library	• Information is displayed in a structured way, depending on the space available in the library • Every section that contains information is accessible by users
	R2: Medium amount of space between objects	• Locate at least one object in a wide 360^{o} displayed environment • Allow movement across fields of information such as 'zoom' or 'pan'
	R3: Provide control over functions	• Allow users to search by typing a keyword • Allow users to measure the length of objects and control their function
(ii) Organization of information	R4: Display information in full	• Display information in a hierarchy and organise data in a database • Manipulate information by using 'pan' and/or 'zoom' to view it
	R5: Display information horizontally	• Information is displayed horizontally
	R6: Display icons/buttons in the same size	• Icons are displayed in a consistent and uniform manner
	R7: Display text in a large, simple font	• Use fonts that are large and simple
(iii) Function	R8: Label functions correctly	• Label objects • Provide tooltips with which to identify objects
	R9: Use terminology which will minimise technical problems	• Include an information bar that does not disturb user activity
	R10: Include animations	• Use a functional animation for icons or images • Use an inspector window to show the current parameters of a moving object
	R11: Provide a help function	• Use a tooltip as a label and to explain the help function • Use a menu to explain the help function • Employ a character as an agent to help to describe the function
(iv) Use of color	R12: Use multiple colors for object which contrast with the color of the background	• Use different colors and tones for different objects
	R13: Use the same color consistently for items which are related	• Use the same range of colors for text and objects which are related to each other to achieve consistency and comprehensibility
	R14: Use bright and natural colors for images/objects	• Use bright colors including yellow, green, orange and turquoise

4 Conclusion

This research explored children's requirements through participatory design activities. This approach was used to analyse a mental model of children's experiences in order to improve their understanding and experience of WebOPAC's interface design, with a specific focus on graphic design. Two participatory activities were conducted which contributed to a list of requirements for better graphic design. IV techniques and VR have the potential to support the implementation of the children's requirements.

References

1. Jochmann-Mannak, H., Huibers, T., Sanders, T.: Children's information retrieval: beyond examining search strategies and interfaces. In: The 2nd BCS-IRSG Symposium on Future Directions in Information Access. British Computer Society, London (2008)
2. Dogget, A.M.: A Statistical Comparison of Three Root Cause Analysis Tools. Journal of Industrial Technology 20(2), 2–9 (2004)
3. Chaparro, B., Baker, J.R., Shaikh, A.D., Hull, S., Brady, L.: Reading Online Text: A Comparison of Four White Space Layouts. Usability News 6(2) (2004), `http://psychology.wichita.edu/surl/usabilitynews/62/whitespace.htm`
4. Neilson, J.: Usability Engineering. Morgan Kaufmann Publishers, San Francisco (1993)
5. Van Barneveld, J.J.F.: User Interfaces for Personalized Information System. Telematica Instituut, pp. 1–82 (2003)
6. Harst, G., Maijers, R.: Effective GI-ontwerp. Schoonhoven. The Netherlands (1999)
7. Galitz, W.O.: The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques. Wiley Computer Publishing, Chichester (2002)
8. Oosterholt, R., Kusano, M., Vries, G.: Interaction design and human factors support in the development of a personal communicator for children. In: Computer Human Interaction. ACM, Vancouver (1996)
9. Benson, C., Elman, A., Nickell, S., Robertson, C.: Z.: Visual Design. GNOME Human Interface Guidelines. T. G. U. Project. Version 1.1. Free Software Foundation, Inc., Boston (2002)
10. Christoffel, M., Schmitt, B.: Accessing Libraries as Easy a Game. Springer, Berlin (2002)
11. Kid's Catalog Web `http://starship.snap.lib.ca.us/`
12. KidsWeb, `http://www.mln.lib.ma.us/kids/`
13. International Children Digital Library, `http://www.icdl.com`
14. Gilutz, S., Nielson, J.: Usability of Website for Children: 70 Design Guidelines, pp. 2–128. Nielson Norman Group, Fremont (2005)
15. Kahn, R., Wilensky, R.: A framework for distributed digital object services. International Journal on Digital Libraries 6(2), 115–123 (2006)
16. Muller, M.J.: PICTIVE-An Exploration in Participatory Design. In: Computer Human Interaction. ACM, Bellcore (1991)
17. Murray, M.: January 26, 2005 07:36:06 PM.: A Closer Look at PICTIVE: A Participatory Design Technique. (2001), `http://www.larson-tech.com/567-Fall2004/pictive.htm` (retrieved June 26, 2006)
18. Baeza-Yates, R., Ribeiro-Neto, B.: User Interface and Visualization. In: Modern Information Retrieval, pp. 258–260. ACM Press, England (1999)

19. Furnas, G.W., Rauch, S.J.: Considerations for Information Environments and the NaviQue Workspace. In: 3rd ACM International Conference on Digital Library, pp. 79–88. ACM Digital Library, Pittsburgh (1998)
20. Johnson, B., Shneiderman, B.: Tree-Maps A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In: Aerospace and Electronics Conference. Dayton: IEEE Conference Proceeding (1991)
21. Beheshti, J., Large, A., Julian, C.A.: Designing A Virtual Reality Interface for Children's Web Portals. In: Graduate School of Library & Information Studies, pp. 1–13. McGill University (2005)
22. Strohecker, C., Barros, B., Slaughter, A.: Mapping Psychological and Virtual Spaces. Online Proceedings of Design Computing on the. International Journal of Design Computing (1998)

Influence of Gender and Age on the Attitudes of Children towards Humanoid Robots

Fang-Wu Tung

Department of Industrial and Commercial Design,
National Taiwan University of Science and Technology,
43,Sec. 4, Keelung Rd.,Taipei,106,Taiwan
fwtung@mail.ntust.edu.tw

Abstract. This study addresses the attitudes of children toward robots displaying various degrees of anthropomorphic appearance. Understanding the means by which children perceive and evaluate robots across the spectrum of anthropomorphism is a crucial issue within the field of robotics research. This study conducted two experiments to understand children's attitudes toward robots with various degrees of realism and examine whether gender or age influences the social and physical attraction children feel toward humanoid robots. The results of the study suggest that when designing robots for children, designers need not focus on creating an authentic human-like appearance. In addition, the influence of children's age on their attitudes toward robots is less significant than that of gender. Generally, children aged from 8 to 14 years have similar attitudes to and perceptions of humanoid robots. An interesting finding is the persistent differences between boys and girls, with respect to the ratings of their social and physical attraction to robots. Particularly, girls are more accepting of human-like robots, especially female robots, than boys are.

Keywords: Human-Robot interaction, child, interaction design, humanoid robot.

1 Introduction

With the advancement of robotic technology, it becomes a tangible future that robots could serve as assistants, companions, or perform other social roles in our environments. Accordingly, the manner people interact with robots is socially situated and multi-faceted. In addition to usability, social and emotional levels of interaction influence a person's acceptance of the roles of robots [1,2,3]. Being an assistant or companion, a robot is expected to communicate with humans in a smooth and natural way, using verbal as well as non-verbal communication. A study by Goetz et al. [4] showed that people anticipate humanlike robots to be best suited for interactive tasks, while mechanical-looking robots were best suited for routine jobs. Thus, robots designed with anthropomorphic characteristics are the embodiment of a human-computer interface, providing the basis for potential social relationships [5,6]. Attributing humanlike characteristics (such as a human appearance) to a robot can facilitate people's understandings of its behaviors or functionalities, leading to form a

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 637–646, 2011.

meaningful social interaction between human and robot. Such notions lead to an assumption that increasing the realism of a robot has practical benefits. While many researchers are exerting efforts to pursue a highly humanlike form for a social robot, we are interested in the degree of realism involved that people deem sociable and appealing. This study specifically addressed the attitudes of children toward robots displaying various degrees of anthropomorphic appearance, as numerous robots have been created to serve as social companions or learning partners for children.

Various robots are specifically designed for children, for educational, entertainment, and therapeutic purposes. Therefore, understanding children's perceptions of humanoid robots with various degrees of realism is essential. Studies related to Computers As Social Actors (CASA) have empirically proven that people tend to treat a computer as a social entity when the computer exhibits social cues adequately to elicit social responses from people [7,8]. This tendency should also apply to the way humans interact with humanoid robots, especially robots that exhibit humanlike appearances. Regarding research that concerns children's attitudes toward robots, the phenomenon of childhood animism should also be considered. Piaget [9] first used animism to describe children's tendency to endow inanimate things with life and consciousness. He outlined four stages of animism through which children normally go. At the first stage (age 4–6), a child attributes consciousness and life to anything that is in any way active, undamaged, or useful. A whole dish is alive; a broken dish is not alive. In the second stage (6－7), only objects that move are given life-like qualities. In the third stage (8－10), consciousness is attributed only to objects that move spontaneously. In the final stage (older than age of 11), a child restricts consciousness and life to plants and animals only. According to Piaget, animism probably demonstrates children's incomplete knowledge and understanding of the world, and changes in the course of development. Okita et al. [10] also revealed that elder children attribute less animistic properties to robots compared to younger ones, but the reductions occur in piecemeal fashion.

As mentioned above, the fact that children over-generalize animacy might lead them to accept a robot and treat it as a social partner, regardless of the degree to which it appears like humans. Children use their animistic intuition to attribute intelligence, biological function, and intention to the objects they encounter [11,12]. This raises the question as to whether robots that interact with children should have a humanoid appearance (an appearance that closely resembles human beings). Understanding the means by which children perceive and evaluate robots across the spectrum of anthropomorphism is a critical issue within the field of robotics research. The phenomenon of childhood animism diminishes gradually as children get older, leading to the prediction that age differences might affect children's perceptions of humanoid robots. In addition to age-related factors, gender differences served as a noteworthy factor in this study as gender differences related to child interaction with agents offered by computers were discovered [13]. Thus, this study further investigated whether age and age differences affect children attitudes toward robots with different degrees of apparent realism.

To summarize, this study aimed at exploring children's attitudes toward humanoid robots with different anthropomorphic appearances. This research was interested in how children differentiate the degrees of realism of robots and examined whether or not these degrees of realism make a difference in their perceptions of humanoid robots regarding social and physical attraction. This study conducted two experiments addressing the research questions listed below.

RQ1. Are children able to differentiate humanoid robots according to the degree of realism?

RQ2. Do children's attitudes toward robots differ with degrees of realism?

RQ3. Do individual differences such as gender and age affect the way children perceive robots?

2 Method

2.1 Experiment1

The first experiment was conducted as a pre-experiment. The experiment investigated children's perceptions and judgment of the degree of human likeness of robots in terms of appearances. To conduct the investigation, a number of robot resources developed by companies, institutes, research labs, and artists were accessed and 54 robots ranging from "barely human" to "fully human" were gathered. One of the 54 robots was gathered from a Japanese movie, in which the robot was played by an actress. Researchers have assumed that the head is the primary place of human-robot interaction. This assumption is empirically proven by research [14] that people's perceptions of humanness of whole robots and robot heads were correlated. The findings indicate that the form of the head plays an important role in the perception of humanness. Thus, this study adopted robot heads as the focus for examining children's perceptions of humanoid robots.

The 54 robot images obtained mostly from the internet were edited by re-scaling and removing variables such as background color, marks, and other objects to ensure that the images were presented in a standardized format. Three professional designers then examined the 54 images to screen out redundant or inappropriate images regarding the anthropomorphism scale and identified 34 images, including that of an actual person, for possible inclusion in the study.

Twenty-nine children aged approximately 10~11 were recruited to sort the 34 images by means of hierarchical clustering individual assessment. Upon beginning a trial, the experimenters shuffled the cards in the set to ensure a random presentation. To facilitate children participating in the study, they were asked to sort the 34 cards into three categories: low, middle, and high degree of realism. Each subset was then divided into three groups based on the same criteria. Thus, each participant sorted the 34 robot images into nine groups ranging from low human likeness to high human likeness. We then used hierarchical clustering methods (Ward's method in SPSS) to analyze the distance matrices for the robots. We obtained four groups to represent various degrees of human-like appearance perceived by the children. Based on ranking data, twelve robots were selected from the four groups to illustrate how children rate robots on an anthropomorphic scale, from "little resemblance to humans" to "highly human-like", (Table1).

Table 1. Twelve images of humanoid robots

											
R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12

(R1: NUVO/ ZMP; R2: Roborior/ Tmsuk; R3:Jonny 5/Lynxmotion; R4:Wakamaru/Mitsubishi; R5:ifbot/ Business Design Laboratory; R6:Nao/Aldebaran Robotics; R7: Nexi M.D.S. Robot /MIT Media Labs Personal Robots Group; R8: Robottina mod.009/Ziopredy; R9:Miss Rong Cheng /Chinese Academy of Sciences; R10: EveR-2 muse / Korean Institute for Industrial Technology; R11: Cybernetic human-HRP/AIST; R12: a character in a Japanese film of “Cyborg She”)

2.2 Experiment2

Experiment 2 was conducted to understand whether or not children’s attitudes toward robots differ with the degrees of realism and to examine the effects of gender and age differences on children’s feelings toward the robots.

Participants. A large sample group of 267 children (N=267) was recruited, including 87 fourth graders (42 girls and 45 boys), 86 sixth graders (42 girls and 44 boys), and 94 eighth graders (45 girls and 49 boys) from five schools.

Measurement Tools. The dependent variables of social attraction and physical attraction were modified from a version of McCroskey and McCain's [15] social and physical attraction scale, as well as from relevant studies that adapted the same scale to measure users’ attitudes toward computers, robots, or media [16,17]. Social attraction and physical attraction are two key dimensions of interpersonal attraction, have been found to be facilitators of interpersonal communication leading to the formation of friendships. Several humanoid robots are being developed to enhance attitudes of social acceptance among users in an attempt to build social relationships. For this reason, this study investigated whether or not different levels of anthropomorphic appearances influence children’s social and physical attraction toward robots.

The social attraction scale consists of five items: 1) “I think this robot is friendly”; 2) “I like this robot”; 3) “I think this robot could be a friend of mine”; 4) “I would like to have a friendly chat with this robot”; and 5) “This robot would be pleasant to be with”. The physical attraction scale consists of three items: 1) “I think this robot is good looking; 2) “I find this robot very attractive physically”; and 3) “I like the way this robot looks”. The two sets were measured using a set of paper-and-pencil questionnaires, each item had a 7-point Likert scale that ranged from 1 (“very strongly disagree”) to 7 (“very strongly agree”). The wording used in the questionnaires was discussed with teachers and the children to prevent any misunderstanding.

Stimuli and Procedure. The stimuli used for this experiment were the twelve robots obtained in Experiment 1. Each robot image was high-quality color printed on a single paper with the aforementioned questionnaires. The participants were asked to evaluate the images of the twelve robots by completing questionnaires, meaning that each participant received twelve sheets to finish the questionnaire survey. The order of the twelve sheets was randomized for each child. The experiment was conducted at select schools. Participants completed the questionnaire survey either in their classrooms or in a quiet place such as a school library.

3 Results

Internal consistency (Cronbach's) was calculated to assess the reliability of these scales. Cronbach's $\alpha$for the social attraction and physical attraction among the 4th, 6th, and 8th graders were all more than 0.7. According to Nunnally [18], Cronbach's α value of 0.7 is adequate for internal consistency reliability. The measures used in the study demonstrated adequate reliability.

Social Attraction. Table 2 shows the mean social attraction scores rated by participants in terms of their grades and genders. Figure 1 illustrates how boys and girls evaluated each robot's social attractiveness. Figure 2 shows the results according to the three grade groups, 4th, 6th, and 8th. The two figures present similar plots illustrating how the positive responses increase with the degree of realism until the point of R7, which thereafter resemble humans closely, but not perfectly. At that point, participants began to react negatively to robots before reaching the point of R11, which thereafter perfectly mimicked human appearance.

The analysis of variance was used to examine the effect of gender and age on the social attraction children felt toward the robots respectively, from "little resemblance to humans" to "highly human-like". Significant gender differences were observed among the rating of images R1, R6, R9, R10, R11, and R12 (all $ps < .01$). The boys were attracted to the image of R1 to a far greater degree than the girls were. With the exception of R1, the remaining images were rated more highly in terms of social attractiveness by girls as compared to boys.

A significant difference was found in the rating of R6 ($ps < .05$). Post hoc analysis indicated that 4th and 6th graders rated R6 higher in terms of social attraction than 8th graders did.

Table 2. Mean social attraction scores rated by participants

	R1	R2	R3	R4	R5	R6	R7	R8	R8	R10	R11	R12
boy	3.7	3.4	3.5	4.4	4.3	5.4	3.7	3.1	2.3	2.3	3.8	4.5
girl	2.9	3.3	3.1	4.1	4.7	5.8	3.8	2.7	3.0	3.0	5.1	5.6
4^{th}	3.6	3.7	3.3	4.6	4.7	5.8	3.7	3.2	2.9	2.9	4.4	4.9
6^{th}	3.2	3.2	3.4	4.4	4.5	5.7	4.0	2.8	2.6	2.6	4.4	5.0
8th	3.2	3.2	3.1	3.9	4.4	5.1	3.5	3.0	2.5	2.6	4.4	5.1

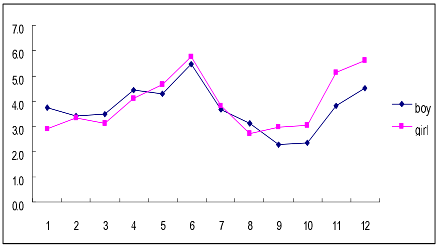


Fig. 1. Each robot's social attractiveness evaluated by boys and girls

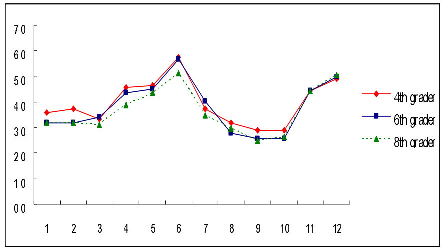


Fig. 2. Each robot's social attractiveness evaluated by different graders

Physical Attraction. Table 2 shows the mean physical-attraction scores rated by participants in terms of their grades and genders. Figures 3–4 illustrate the physical attraction children felt toward the twelve stimuli. Similar preferences for the twelve robots were observed regarding physical attractiveness. An ANOVA was performed to examine whether age or gender influenced children's attraction to the appearance of the robots in the twelve images. Significant gender differences were observed among the ratings for R1, R5, R9, R10, R11, and R12 (all ps$<$.05). A similar pattern was observed in which boys liked the appearance of R1 far more than girls did, and girls liked the appearances of the other images more than the boys did.

Results show no significant age differences in children's physical attraction toward the twelve robots.

Table 3. Mean Physical attraction scores rated by participants

	R1	R2	R3	R4	R5	R6	R7	R8	R8	R10	R11	R12
boy	3.3	2.9	3.0	3.8	3.4	4.7	3.0	2.8	2.1	2.3	3.7	4.4
girl	2.7	3.0	2.6	3.7	4.0	5.4	3.2	2.5	2.6	2.8	5.1	5.6
4^{th}	3.0	3.2	2.9	4.0	4.0	5.2	3.2	3.0	2.6	2.8	4.4	4.7
6^{th}	3.1	2.9	3.0	3.7	3.5	5.3	3.0	2.5	2.3	2.4	4.3	4.8
8^{th}	2.9	2.8	2.6	3.5	3.7	4.7	3.0	2.4	2.2	2.5	4.4	5.3

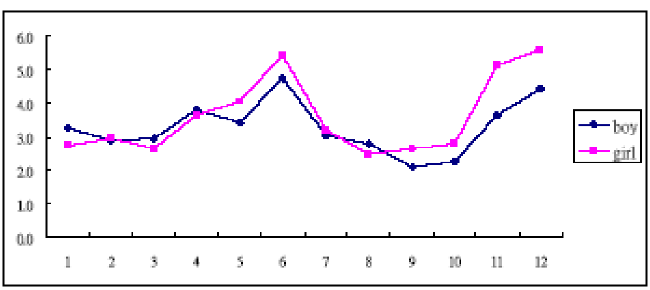


Fig. 3. Each robot's physical attractiveness evaluated by boys and girls

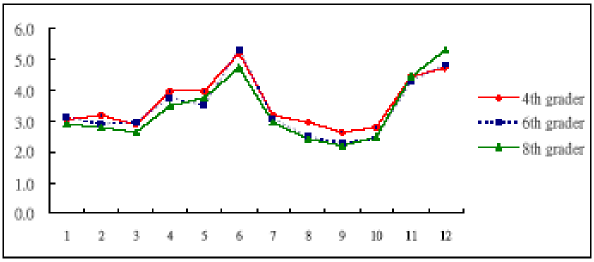


Fig. 4. Each robot's physical attractiveness evaluated by different graders

4 Discussion

The results from the current study show that the degrees of realism of robots make a difference in children's attitudes towards robots, echoing the theory of the uncanny valley, proposed by Mori [19]. Figures1– 4 show a continuously positive change of realism versus children's evaluation of social and physical attraction until a point of realism beyond which children's evaluation abruptly decreases. As the appearance becomes less distinguishable from a human, children's responses become positive once again. An uncanny valley was found in this study, indicating that children were less attracted to the images considered highly human-like but still distinguishable from humans, thereby evoking a feeling of discomfort in the observer. The study empirically supports related research that has argued that robots designed for children should not focus on creating a completely human-like appearance, unless they are perfect replicas and indistinguishable from humans. Results suggest that R6 and R12 received high evaluation from children in terms of social and physical attractiveness. Image R12 is a real person appearing young and pretty, while R6 is rated as a middle degree of realism. A striking finding is that the top of the first peak of the curve is slightly higher than the point at which R12 is located, differing from the results of Mori's uncanny valley where the highest score falls on the real human. This study attempted to identify the threshold of humanness that is appropriate for robots designed for children. The implication of the observation is that designing robotic appearances for children might consider combining human and machine features to restrict effort towards the first peak of the uncanny valley, rather than attempting to replicate human appearance perfectly.

The finding that children prefer R6 most among the stimuli is in line with Woods's [20] finding that children prefer robots with cartoon-like appearances. Humans, including children, are sensitive to the particular pattern of features that form a face. Using mere representations of cartoon-like faces can avoid the uncanny valley and cover a large aesthetic range [21]. Completely eschewing a human likeness is not necessary to avoid the uncanny valley. Uncanny faces can be modified with designs to increase acceptance by emphasizing features indentified with friendliness and youthfulness [21]. In terms of the robotic face design, DiSalvo et al. [23] suggested that head length and width influence perceptions of a robot's human likeness. They suggested that the head should be slightly wider than it is tall and the eye space should be slightly wider than the diameter of the eye. Designers might consider exaggerated features and an encasement to hide mechanical parts to ensure that the robot appears not only humanlike but also product-like. Among the twelve robot images, the proportion of R6 most closely conformed to the aforementioned design guidelines, which could account for why R6 received the highest evaluation regarding social and physical attractiveness.

To examine the individual differences such as age and gender, the results show that the influence of children's age on their attitudes toward robots is less significant than that of gender. The results indicate that children, including 4th, 6th, and 8th graders, have similar attitudes to and perceptions of humanoid robots. Those children aged from 8 to 14 years are in the third and fourth stages according to Piaget's theory of cognitive development. No age-related factors observed in this study imply how children's perception of digital objects or computational companions might be evolving because they were born in a digital era. This is a question remaining to be answered.

In contrast to no age-related factors found, an interesting observation involved the persistent differences between boys and girls with respect to the ratings of their social and physical attraction to robots. The differences were found in their attitudes towards robots with the lowest degree of human-likeness and a high degree. Boys reported higher social and physical attractiveness of the R1 robot, which is evaluated as the lowest degree of human likeness among the stimuli. A possible explanation could be that boys are usually more familiar with mechanical tools than girls are and boys were more likely to consider a robot with a mechanical appearance attractive. However, the preference changed while analyzing attitudes toward robots with high anthropomorphic appearances. Results show that girls felt more social and physical attraction to those high human-like robots than did boys. A study by Green et al. [24] indicated that women have greater tolerance in the acceptable range of facial proportion of people and robots. They concluded that traditional feminine nurturing roles because increased acceptance and males are generally more familiar with mechanical objects with a diminished need to anthropomorphize them. Greater female sensitivity to nonverbal cues such as facial expressions may be the result of socialization, gender patterns in emotional learning, or gender differences in the brain [25,26]. Interestingly, gender differences in the perception of facial expressions in the real world also influence the way children interact with humanoid robots. Another possible explanation for the tendency among girl participants to favor those humanoid robots with high degrees of anthropomorphism is that those robots are presented as being female. The findings echo some studies that females favor their own gender while males have no gender preference [27,28,29].

To summarize, this study observed the way children differentiate the degrees of realism of humanoid robots and obtained their attitudes toward robots with different anthropomorphic appearances. Findings regarding age and gender differences in their perceptions of humanoid robots were discussed. The results of this study could provide valuable references for designers and manufacturers of robots.

Acknowledgments. This material is based upon work supported by the National Science Council of the Republic of China under grant NSC98-2221-E-239-012-MY2.

References

1. Bartneck, C., Forlizzi, J.: A design-centred framework for social human–robot interaction. In: Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication, pp. 591–594. IEEE Press, Kurashiki (2004)
2. Kiesler, S., Hinds, P.: Introduction to This Special Issue on Human-Robot Interaction. In: Human- Computer Interaction, vol. 19, pp. 1–8 (2004)
3. Young, J., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H.I., Grinter, R.E.: Evaluating Human-Robot Interaction: Focusing on the Holistic Interaction Experience. International Journal of Social Robotics 3(1), 53–67 (2011)
4. Goetz, J., Kiesler, S., Powers, A.: Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: Proceedings of the Twelfth IEEE International Workshop on Robot and Human Interactive Communication, Lisbon, Portugal (2003)
5. Duffy, B.R.: Anthropomorphism and the Social Robot, Special Issue on Socially Interactive Robots. Robotics and Autonomous Systems 42(3-4), 31 (2003)
6. Kanda, T., Ishiguro, H.: An approach for a social robot to understand human relationships-Friendship estimation through interaction with robots. Interaction Studies 7(3), 369–403 (2006)
7. Nass, C., Steuer, J.: Anthropomorphism, Agency, and Ethopoeia: Computers as Social Actors. Human Communication Research 19(4), 504–527 (1993)
8. Reeves, B., Nass, C.: Media equation: how people treat computer, television, and new media like real people and places. Cambridge University Press, New York (1996)
9. Piaget, J.: The child's conception of the world. Littlefield Adams, Savage (1929)
10. Okita, S.Y., Schwartz, D.L.: Young Children's Understanding of Animacy Entertainment Robots. International Journal of Humanoid Robotics 3(3), 393–412 (2006)
11. Carey, S.: Conceptual Change in Childhood. MIT Press, Cambridge (1987)
12. Okita, S., Schwartz, D., Shibata, T., Nakamura, O., Tokuda, H.: Exploring Young Children's Attributions through Entertainment Robots. In: 14th IEEE International Workshop on Robot and Human Interactive Communication, Nashville, TN (2005)
13. Hall, B., Henningsen, D.: Social facilitation and human–computer interaction. Computers in Human Behavior 24, 2965–2971 (2008)
14. DiSalvo, C., Gemperle, F., Forlizzi, J., Kiesler, S.: All robots are not created equal: The design and perception of humanoid robot heads. In: Proceedings of DIS 2002, pp. 322–323. ACM Press, New York (2002)
15. McCroskey, J.C., McCain, T.A.: The measurement of interpersonal attraction. Speech Monographs 41, 261–266 (1974)
16. Moon, Y.: Similarity effects in human-computer interaction: Effects of user personality, computer personality, and user control on attraction and attributions of responsibility, Unpublished doctorial dissertation, Stanford University, Stanford, CA (1996)
17. Lee, K.M., Peng, W., Yan, C., Jin, S.: Can Robots Manifest Personality? An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction. Journal of Communication 56, 754–772 (2006)
18. Nunnally, J.C.: Psychometric Theory. McGraw Hill, New York (1967)
19. Mori, M.: The uncanny valley. Energy 7(4), 33–35 (1970)
20. Woods, S.: Exploring the Design Space of Robots: Children's Perspectives. Interacting with Computers 18, 1390–1418 (2006)
21. Blow, M., Dautenhaun, K., Appleby, A., Nehaniv, C.L., Lee, D.: The art of designing robot faces: Dimensions for human-robot interaction. In: Proceedings of HRI 2006, pp. 331–332. ACM Press, New York (2006)

22. Hanson, D.: Exploring the aesthetic range of humanoid robots (electronic resource). In: Proceedings of the ICCS/CogSci2006 Long Symposium: Toward Social Mechanisms of Android Science, pp. 16–20. Cognitive Science Society, Vancouver (2006)
23. DiSalvo, C., Gemperle, F., Forlizzi, J., Kiesler, S.: All robots are not created equal: The design and perception of humanoid robot heads. In: Proceedings of DIS 2002, pp. 321–326. ACM Press, New York (2002)
24. Green, R.D., MacDorman, K.F., Ho, C.C., Vasudevan, S.K.: Sensitivity to the proportions of faces that vary in human likeness. Computers in Human Behavior 24(5), 2456–2474 (2008)
25. Hall, J.A., Matsumoto, D.: Gender differences in judgments of multiple emotions from facial expressions. Emotion 4(2), 201–206 (2004)
26. Lytton, H., Romney, D.M.: Parents' differential socialization of boys and girls: A meta-analysis. Psychological Bulletin 109(2), 267–296 (1991)
27. Nosek, B.A., Banaji, M.R.: The go/no-go association task. Social Cognition 19(6), 161–176 (2001)
28. Rudman, L.A., Goodwin, S.A.: Gender differences in automatic in-group bias: Why do women like women more than men like men? Journal of Personality and Social Psychology 87, 494–509 (2004)
29. MacDorman, K., Coram, J., Ho, C., Patel, H.: Gender differences in the impact of presentational factors in human character animation on decisions in ethical dilemmas. Presence 19(3), 213–229 (2010)

Part VII

Playing Experience

Affective Videogames: The Problem of Wearability and Comfort

Andrea Bonarini[1], Fiammetta Costa[2], Maurizio Garbarino[1], Matteo Matteucci[1], Maximiliano Romero[2], and Simone Tognetti[1]

[1] Department of Electronics and Information, Politecnico di Milano
[2] Health Care Design, Industrial Design Department (INDACO), Politecnico di Milano
maximiliano.romero@polimi.it

Abstract. The aim of this paper is to report the interaction design process followed by an interdisciplinary team to develop an innovative ICT wearable device for affective video gaming. The process follows Norman and Draper's User Centered Design principles [1] including: functional development, laboratory test of the technology with human subjects, product design, prototype realization and experimentation with final users. The functioning of the device is based on the detection of physiological parameters, e.g., Blood Volume Pulse (BVP), Temperature (T), and Galvanic Skin Response (GSR), through electrodes placed on the forehead of the player. These signals are aimed at detecting the emotional state of the player by means of computational intelligence algorithms. This information can be used to modify the behavior of a videogame in order to maintain the player in the desired state of subjective enjoyment. Our goal was to develop a comfortable and easy to use device to avoid disturbs on the emotional state of the player.

Keywords: Design for wearability, Affective computing, Videogame.

1 Introduction

Videogames have seen recently a paradigm shift in user interaction; from passive devices such as joypad and keyboards they have moved to devices able to directly capture the player's intention from body movements and gesture [2]. This shift has shown how a strong relationship between game and user could be established through the use of devices improving the interaction. This aspect has often turned to be more important than perceptual realism of the game. To push this relationship further, we need to adapt the game to each single user monitoring her/his engagement and acting on the game parameters to keep her/his engaged all the time.

In this paper, we present a device to capture data to infer the user mood while interacting with a videogame. This device leverages on recent studies in affective computing that present computational intelligence algorithm for the detection of user preferences [3] and emotional state [4] from physiological signals such as Blood Volume Pulse (BVP), Temperature (T), and Galvanic Skin Response (GSR). Since

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 649–658, 2011.

the emotional state of the player may be affected by the presence of an invasive sensing device, we aimed at developing a comfortable and easy-to-use device.

The present work has been developed as collaboration between two research teams from the department of Electronics and Information and the department of Industrial Design of Politecnico di Milano. The first group, working in the Artificial Intelligence and Robotics Laboratory (AIRLab), has a background in electronic engineering and computational intelligence, while the Health Care Design (HCD) team is specialized in technological and user requirement integration.

While the design of the electronic back-end of the device is out of the scope of this paper, we focus here on the development process of the front-end hardware. This process had to face three main challenges; it should be easy to wear, it should not affect the interaction with the game, and it should provide reliable and artifact-free signals.

2 Process Description and Motivations

The technological feasibility of emotional detection from physiological signals in video gaming has been recently verified experimentally [3, 4]. The test, done on 75 volunteers, has shown how it is possible to establish a reliable relationship between physiological data and preference of the subject for a specific setting of the game with respect to alternative ones. The game used in the experiment was the TORCS car racing game [5], and the racing settings that have been compared were affecting the opponent's skills.

During the experiment, we have acquired a large number of physiological signals, and we have identified by statistical analysis those providing the larger amount of information needed to classify the emotional state. These have been acquired with a traditional data acquisition system, which requires the subject to be instrumented with sensors on fingers and chest, and these sensors are then wired to an acquisition device. However, by observing the subjects, we have noticed how the instrumentation made many subjects uneasy, possibly affecting both their performance and their emotional relationship with the game. Therefore, we have decided to design a new wearable interface that could be easily accepted by any player, and, at the same time, could provide enough information for emotion mapping.

Being both hands involved in the racing game (but this applies to other videogames as well) we have identified the forehead as a good candidate place to get the most interesting signals such as BVP, T and GSR. An acquisition system worn on the head makes it also possible to get signals from accelerometers and gyroscopes to monitor head movements, that can be useful to complement the physiological signals. The developed headset has been connected to the acquisition system via wireless communication.

On the market, there are some headsets with electrodes that have also been targeted to the videogame market (e.g., NIA by OCZ, Mindset by Neurosky, EPOC by Emotiv). However, they are mainly aimed at controlling the game through electric surface potentials while they are not aimed at detecting features that could be related to the affective state of the user such as BVP or GSR.

To fill this market gap, three incremental steps have been performed regarding product design: the first step was to package the electronic components in a compact product, the second was to increase the wearability considering anthropometrical data [6] and pressure zones; the third concerns user acceptance considering aesthetic characteristics.

User comfort is not only related to the physiological condition, but psychological factors have also to be taken into account. In our case, this is particularly relevant because the goal is specifically to deduce psychological information rather than collect physiological data. Furthermore, the formal design of the packaging is important also for user acceptance since it should evoke the game world and not clinical instrumentation like most of the data acquisition devices available on the market. Several demonstrators have been produced with rapid prototyping techniques to test wearability and pleasure [7] with final users and achieve a satisfactory release.

3 Design Process

The AIRLab team initiated the present research but when the back-end has been sufficiently developed, it was necessary to focus on front end of the system. For this reason, the HCD group has been involved.

HCD group usually collaborate in multidisciplinary teams with ICT and Health Care experts and has developed a specific design process method that takes into account final users participation and co-design [8]. The aim of this method is to involve final users on the front-end side of the technological products from the early stage of the project in order to transfer User Experience to the back-end developers. Actually, if we consider a device as a complex product composed by Hardware and Software, final users might interact with both but they face mainly the front-end side. The back-end side involves engineering aspects and, normally, users are not aware of this layer.

This project was specifically focused on wearability and its impact on biosignals quality. Wearability is essential for a pleasant user experience of an affective videogame and it is also directly related to quality of biosignal acquisition depending on the sensors position and proper skin contact.

The back-end hardware is composed by two sensors that must be directly in contact with the skin, 2 Printed Circuit Board (PCB) and 1 battery. The initial effort was to reduce weight and dimensions and the second step was to decide the location of every component. According to Gemperle study [9], one of the main criteria for determining placement for dynamic wearability is to choose areas that have low movement and flexibility even when the body is in motion. We chose to locate the device on the head for two main reasons: all the needed signals could be collected and the head motion during videogaming was relatively low.

At the beginning, the internal tests have been realized with the prototype 1 (**Fig. 1a.** top left), when design team started to collaborate we developed the second prototype (**Fig. 1b.** top right) with rapid prototyping technique. After this, the back-end hardware has been miniaturized and the third prototype has been developed (**Fig. 1c.** bottom left). At the end of the process has been developed the prototype 4, applied for the user test (**Fig. 1d.** bottom right)

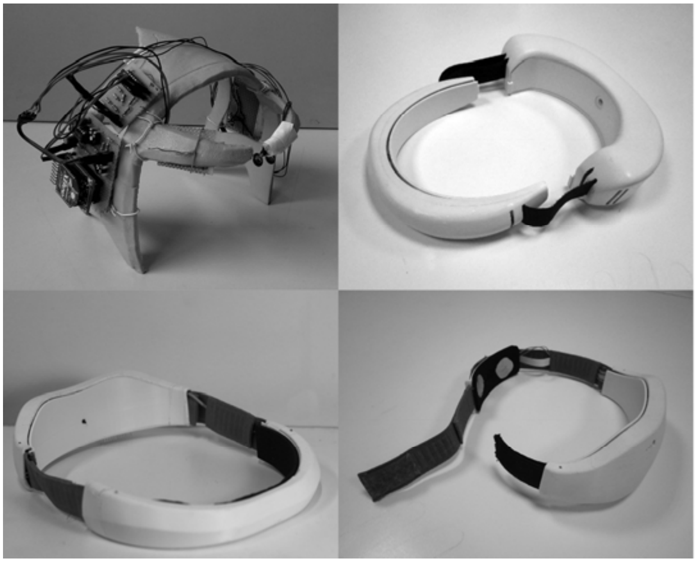


Fig. 1. Prototype evolution: (a) first prototype for laboratory tests, (b) designers' concept, (c) PCB miniaturization, (d) sensors contact improvements

Internal dimensions of front and back cases have been decided on the basis of literature anthropometrical data. Nevertheless, the variability of the head dimensions suggested to use fewer hard shapes and more elastic bands in order to allow the frontal case to become more flexible and improve the contact between biosensors and forehead.

4 Instruments and Method

The final test was performed in laboratory, by involving end users. We used the last prototype of our wearable device and a reference device, the Procomp Infiniti, for comparison. (**Fig. 2.**)

Our goal was to test wearability characteristics and quality of biosignals captured. For this reason, we decided a test protocol in which users were asked to perform different actions and, for each of them, we collected quantitative information. Users were not asked to play the videogame as we preferred them to concentrate on the feeling of wearability. The set up of the test included an empty desk with two separated boxes 80 cm far from the user. On the desk in front of the user our device and the reference device were located. One researcher has checked data transmission between device and computer and has taken note of time of performance, while the second has guided the users on the test.

During the test, users have been asked to dress the two devices (first the wearable device and then the reference one on the left hand), to stand still in a precise position (with both hands on the desk) and to perform the following series of movements:

Fig. 2. Devices used for the final comparison test with user involvement: Procomp Infiniti hands sensors (left) and our last prototype of headset (right)

1. To reach the left box with right hand with natural movement (random motor strategy) and come back to neutral position.
2. To reach the right box with the left hand (sensorized hand) with natural movement (random motor strategy) and come back to neutral position.
3. To reach contemporaneously the right box with right hand and the left box with left hand, then come back to neutral position.

Finally, we ask them to remove the two devices (first the reference device and then the wearable one). At the end of the test we asked each user to fill a questionnaire with the following five questions:

a. Which device is easier to wear?
b. Which device is more immediate for the proper placement of the electrodes?
c. Which device is more immediate for the closure system?
d. Which device allows for more natural gestures?
e. Which device generates more sweat?

5 Results

We first present the analysis of questionnaires that qualitatively measure the wearablility of the devices from the user perspective; then we present a quantitative analysis of placement and removal times and a quantitative analysis of signal quality during both rest and movement periods.

5.1 Questionnaire Analysis

The questionnaire answer expresses the users preference, with respect to a given feature, between the wearable and reference device.

In Figure 3, histograms of the answers are reported for each question. All subjects reported that–our device can be worn more easily than the reference device(a). Less remarkable differences have been observed on the easiness by which people have understood how the device has to be positioned correctly; one user did not understand that the wearable device had to be placed on the head)(b). Users have perceived the closure system of the two devices equivalently complex(c); note that one of them reported a problem with our device related to the volume of the hair that made very difficult the closure of the system. All subjects agreed on the optimality of our device to allow natural movements (d). Finallly, most of the users reported that our device tended to increase the sweat activity(e).

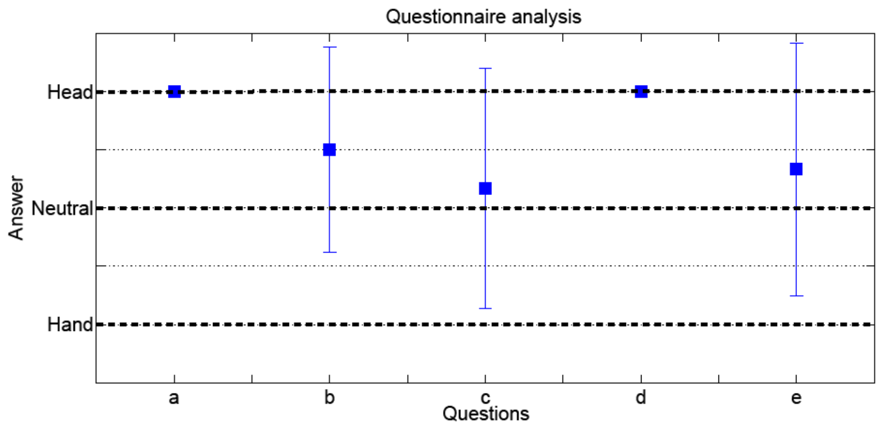


Fig. 3. Users' answers to usability and wearability questionnaire

Questions (a) and (d) confirm that our device we designed, meets the two fundamental objectives for an affective computing device of being easy-to-wear and to not compromise the standard interaction. Question (b) indicates that it is not so easy to understand how to correctly wear the headset. Question (c) indicates that the closure system has to be improved and finally question (e) indicates that our headset produces the side effect of increasing the sweat activity.

5.2 Placement and Removal Times

We measured the time required by a subject to correctly wear and unwear the devices. Each subject has been first shown how each device should be correctly placed. Note that both the devices have an equivalent number of physiological sensors. We are not considering the accelerometer that is present in our device while it is not available in the reference device. In Figure 4, the mean value and its confidence interval (with alpha=0.05) of placement and removal times for each devices are reported. Placement times are significantly higher for the reference device, positioned on the hand, compared to our wearable device (i.e., 12.58 s for the wearable device and 42.75 s for the reference one). A similar consideration holds for the removal times that have a mean value of 4.41 s of our device and 14.98 s the reference one.

Considering the assumption that the lower the placement/removal time is and the higher the easiness to wear is, this result confirms that we have satisfied the objective of building a device that is easy-to-wear.

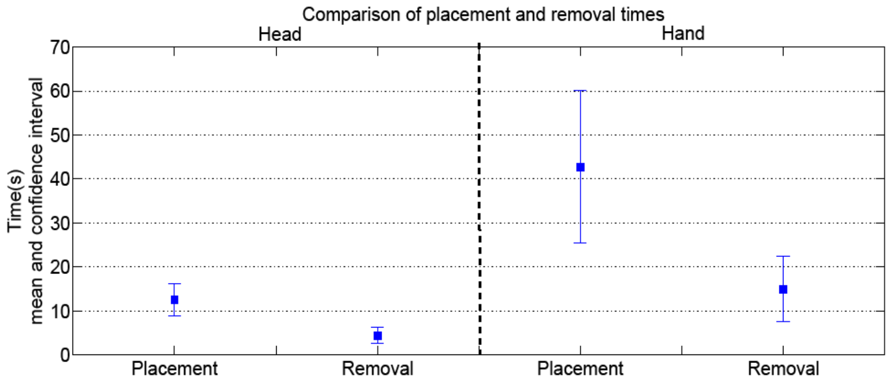


Fig. 4. Placement and removal times are significantly lower for our headset

5.3 Signal Quality

The last part of presented results concerns the quality of the signals obtained from our wearable device. Each wearable device, becoming more wearable tends to lower the signal quality since data are measured in suboptimal positions. However, even small movements could reduce the quality of the signal acquired by a standard medical device. Thus, in a natural environment we could accept to reduce the quality of signals by allowing the measurement even during movements.

In Figure 5, a comparison of BVP signals acquired by both devices during a resting period is presented. The signal from the forehead preserves almost all the intrinsic characteristics (i.e., waveform and positions of heart beats) with respect to a standard signal measured on the finger. A deeper analysis is required to understand whether the upper and lower envelop of BVP, representing the systolic and diastolic pressure, are also preserved.

In Figure 6 is presented a comparison of heart rate signals extracted from both segments of Figure 5. Note that the heart rate signals are very similar: there are only minimal differences that are due to the different position of detected heart beats. Such small differences do not decrease performances of an affective classifier.

In Figure 7 is presented a comparison of BVP signals acquired by both devices during a movement. In this situation, only the headset was able to measure continuously the signal without degradation. The reference device, instead, produced a signal that is not usable. The standard BVP signal is corrupted by the movement artifacts that, if not detected and removed, could lower the performances of an affective classifier [4].

To give an exhaustive analysis of BVP quality during the movements, we have manually evaluated each signal portion corresponding to a movement. We assigned them a quality value according to the percentage of signal that contains valid heart beats (i.e., the typical waveform is preserved). For instance, in the situation reported in Figure 7, our device has 100% quality, while the reference device has 0% quality. In Figure 8, the results of this analysis are reported. In particular for each movement (A) right hand, (B) left hand, (C) both hands, the mean value and confidence interval (with alpha=0.05) of signal quality are presented for both devices. We recall that signals have been acquired in parallel by both devices during each movement. Figure 8 shows that our device, placed on the head, measures the BVP signal with an overall quality of 80% during all the movements while the standard device, placed on the hand, produces a BVP with an overall quality of 20% during all the movements. It is interesting to notice that movement C produces a quality slightly higher than A and B. This is due to the fact that in movement A and B there is also head torsion.

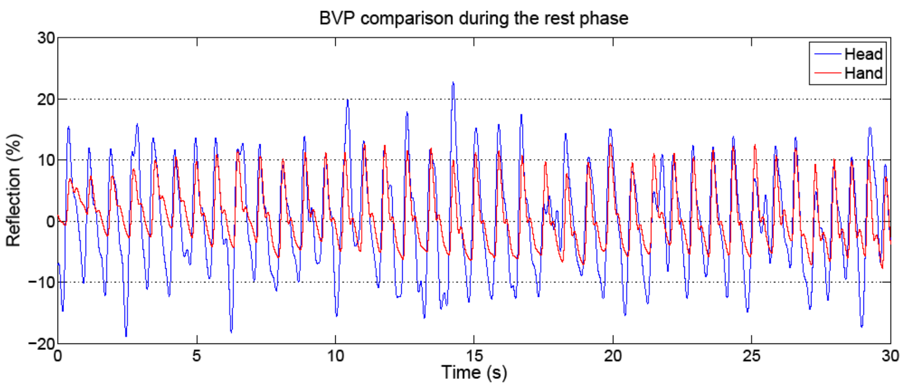


Fig. 5. Comparison of BVP signal acquired from finger and from forehead

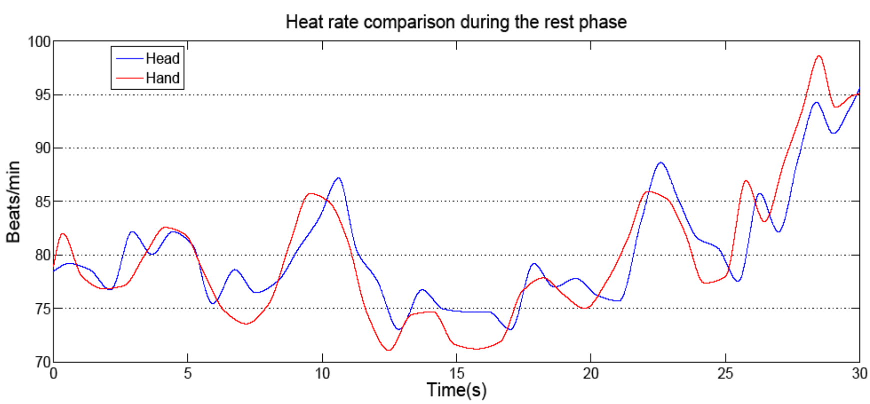


Fig. 6. Comparison of heart rates signal acquired from finger and from forehead

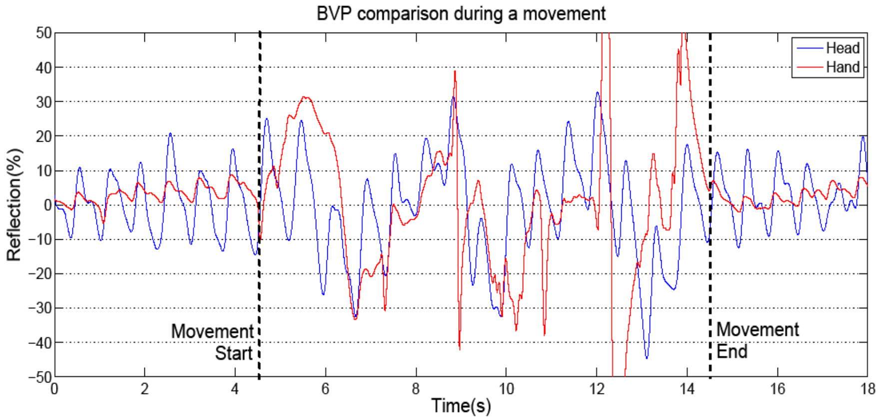


Fig. 7. Comparison of BVP signals during movements

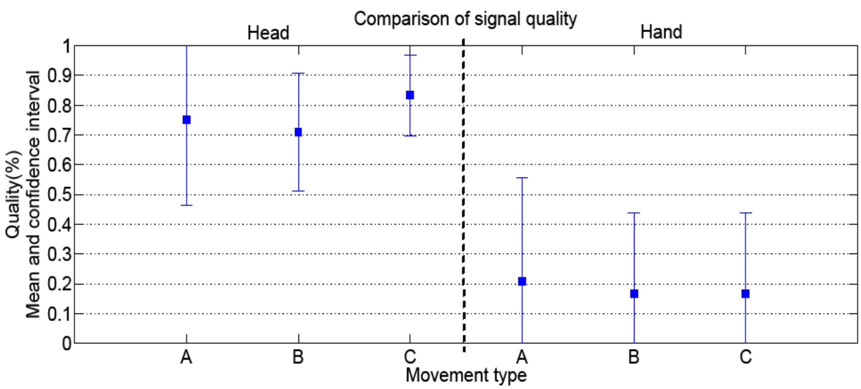


Fig. 8. Quality of signal during movements. Right hand (A), left hand (B), both hands (C)

This result demonstrated that the wearable device we designed acquires the BVP signal with a good quality during the movements that usually characterize activities in natural environments.

6 Conclusion

In this paper we presented the interaction design process followed by an interdisciplinary team to develop an innovative ICT wearable device for affective video gaming. The objective of this work was to design a wearable device that has the following requirements: it should be easy to wear, it should not affect the interaction with the game, and it should provide reliable and artifact-free signals.

The interaction design process we followed allowed us to better satisfy the device requirements. From one side, the HCD group, working of the front-end, aimed at satisfying the wearability aspects while the Artificial intellingence group, working on the back-end aimed at satisfying the signal quality aspects.

The results we obtained demonstrate that the device requirement have been fairly satisfied. The questionnaires analysis shows that, from the user perspective, the wearable device is easy to wear and does not compromise the interaction. The placement/removal times analysis confirms that the wearability aspect has been met. Finally the signals analysis demonstrates that the wearable device produces better signals during a movement compared to a reference medical device.

Acknowledgments. We would like to thanks Paolo Villa, that designed the electronic hardware within his master thesis at Politecnico di Milano. We also thanks to Maicol Zoia, Marco Tonet and Davide Veronesi that help us to prototype the device.

References

1. Norman, D.A., Draper, S.W.: User centered system design: New perspectives on human-computer interaction. Lawrence Erlbaum Associates, New Jersey (1986)
2. McArthur, V., Castellucci, S.J., MacKenzie, I.S.: An empirical comparison of "Wiimode" gun attachments for pointing tasks. In: Proceedings of the ACM Symposium on Engineering Interactive Computing Systems - EICS 2009, pp. 203–208. ACM, New York (2009)
3. Tognetti, S., Garbarino, M., Bonarini, A., Matteucci, M.: Modeling player enjoyment from physiological responses in a car racing game. In: Proceedings of IEEE International Conference on Computational Intelligence and Games (CIG), pp. 321–328. IEEE Computer Press, Los Alamitos (2010)
4. Tognetti, S., Garbarino, M., Bonanno, A.T., Matteucci, M., Bonarini, A.: Enjoyment Recognition From Physiological Data in a Car Racing Game. In: Proceedings of 3rd International Workshop on Affective Interaction in Natural Environments (AFFINE), pp. 1–6
5. The open racing car simulator website, `http://torcs.sourceforge.net/`
6. Peebles, L., Norris, B.: Adultdata: the handbook of adult anthropometric and strength measurements: data for design safety. Department of Trade and Industry, Government Consumer Safety Research, p. 294, London (1998)
7. Jordan, P.W.: Pleasure with Products: Beyond Usability. Taylor and Francis, Abington (2000)
8. Romero, M., Perego, P., Andreoni, G., Costa, F.: New Strategies for Technology Products Development in Health Care. In: MengJooEr (ed.) New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems, Sciyo (2010); ISBN: 978-953-307-213-5
9. Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., Martin, R.: Design for wearability. In: Second International Symposium on Wearable Computers (Cat. No.98EX215), pp. 116–122., doi:10.1109/ISWC.1998.729537

Extraversion and Computer Game Play: Who Plays What Games?

Xiaowen Fang and Miaoqi Zhu

School of Computing, College of Computing and Digital Media
DePaul University
xfang@cdm.depaul.edu, miaoqi.zhu@gmail.com

Abstract. In the last two years, our research group has been conducting online surveys on enjoyment of computer game play and player personalities. From the preliminary data we collected as of now, we observed positive relationship between the personality trait, extraversion, and players' choice of games. This paper presents a research project that attempts to explore this relationship and investigate the reasons why this relationship exists using interview method. The expected findings will likely help game designers improve interaction designs in computer games.

Keywords: Extraversion, personality, enjoyment, computer games.

1 Introduction

The popularity of digital (computer and video) games has reached phenomenal proportions. Based on the statistics provided by Entertainment Software Association ([1]), 1.U.S. computer and video game software sales generated $10.5 billion in 2009, 67 % of American households play computer or video, and their worldwide markets are expected to grow strongly also in the future. Computer games have become a major form of entertainment. In addition, digital games are used increasingly for therapeutic, educational, and work- related purposes ([2], [3]). Given the prominence of computer games for entertainment, researchers need to acquire a better understanding about computer game players and their play experience. However, as Bateman and Boon ([4]) stated in the preface of their book, "a certain mystery still surrounds game design, and although much has been written on the subject, the formal study of game design practices in a definite sense is still in its infancy."

Based on a conceptual framework of computer game play proposed by Fang, Chan, and Nair ([5]), this study investigates the relationship between the personality trait, extraversion, and players' choice of games. The key research questions are: 1) what types of people play what games? And 2) Does the personality trait, extraversion, matter in players' choice of games? We expect that findings of these important questions will no doubt help game designers improve interaction designs in computer games.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 659–667, 2011.

2 Background Literature

Much of the psychological research on games has been focusing on negative effects of violent video games although some recent studies have changed this tone and started to investigate individual differences among game players. In this section, we first review prior research on personality and computer game play. Then we examine the Big-Five personality model and its personality traits.

2.1 Personality and Computer Game Play

Previous research has consistently shown that exposure to violent video games is significantly linked to increases in aggressive behavior, aggressive cognition, aggressive affect, and ardiovascular arousal, and to decreases in helping behavior ([6]). Anderson and Dill ([6]) also suggest that the positive relationship between violent video game play and aggressive behavior and delinquency is stronger for individuals who are characteristically aggressive and for men. Furthermore, a few other studies show that personality is linked to gaming behaviors. Fetchenhauer and Huang ([7]) indicate that the justice sensitivity could be used to predict decisions in a number of games using theoretical paradigms (dictator games, ultimatum games, and a combination of these two games). Douse and McManus ([8]) suggest that players of a fantasy Play-By-Mail game were less feminine, less androgynous, and more introverted than matched controls. The fantasy game players showed lower scores on the scale of empathic concern, and were more likely to describe themselves as "scientific," and to include "playing with computers" and "reading" among their leisure interests than players in the control group. In a more recent study, Whang and Chang ([9]) explored the lifestyles of online game players. Based on an online survey, they classify lifestyles of game players into three groups: single-oriented players, community-oriented players, and off-real world players. Players in each group display distinct differences in their values and game activities, as well as in their anti-social behavior tendencies. This study further suggests that differences in game players' lifestyles reflect not only their personality but also their socio-economic status within the virtual world constructed through game activities.

Bateman and Boon ([4]) applied to Myers-Briggs Type Indicator ([10]) to games and conducted a survey to investigate game player types. The survey had two components: a 32-question Myers-Briggs personality test and a short questionnaire to determine elements such as game purchasing and playing habits. About four hundred participants took part in the study. Based on cluster analysis results, Bateman and Boon ([4]) identified the following four play styles:

- Type 1 Conqueror play involves winning and beating the game.
- Type 2 Manager play revolves around a strategic or tactical challenges.
- Type 3 Wanderer play in which players search for a fun experience.
- Type 4 Participant play.

Within each of these four types, players were further categorized into two subtypes: hardcore and casual players.

Bartle ([11]) recognizes four types of game players who play games in the virtual world: 1) Achievers like acting on the virtual world. Their aim is usually to succeed in the context of the virtual world. 2) Explorers like interacting with the virtual world. They act in order to find out things about the virtual world and how it works. 3) Socialisers like interacting with other players. They like talking, being part of a group, and helping others. 4) Killers like acting on other players. Sometimes, this is to gain a big bad reputation, but other times it's to gain a big good reputation.

More recently, Fang and Zhao ([12]) find: 1) Sensation seeking has a significant and positive effect on enjoyment of computer game play through enhanced engagement during game play for action/ adventure/shooting/fighting, role playing, and sport/racing games. 2) Sensation seeking has a significant and positive effect on enjoyment of computer game play through enhanced cognition values for family entertainment/simulation games. 3) Self-forgetfulness has a significant and positive effect on enjoyment of computer game play through enhanced engagement during game play for role playing games.

Despite some of the groundbreaking work in prior studies, few research has systematically examined what role player's personality plays in choosing games to play. This paper reports the first attempt to address this research question.

2.2 The Big-Five Personality Model

Personality can be defined as a stable set of tendencies and characteristics that determine the commonalities and differences in people's psychological behavior (thoughts, feelings and actions) that have continuity in time. Personality is one of the most elusive areas of psychology, difficult to understand, and difficult to test. Nevertheless, psychologists have developed several theories to explain personality based on two principles: core of personality and periphery of personality. Core of personality addresses the inherent attributes of human beings which do not change over the course of living. They are used to explain the similarities among people. Periphery of personality, on the other hand, focuses on learned attributes. It helps to identify the differences among people.

Over the years, the big five-factor personality model ([13], [14], [15], [16]) has gained acceptance among researchers because it establishes a common taxonomy (Goldberg, 1990). It contains the following five dimensions (or traits) of personality:

- Extraversion: this factor has been the largest. It contrasts traits such as talkativeness, liveliness, and outgoingness versus shyness, quietness, and passivity.
- Agreeableness: the second factor. It contrasts traits such as kindness and gentleness with rudeness and harshness.
- Conscientiousness: the third factor. It includes traits such as organization, discipline, and thoroughness versus sloppiness, laziness, and unreliability.
- Emotional stability (versus neuroticism): the fourth factor. This factor contains traits such as relaxedness, versus moodiness, anxiety, and touchiness.
- Intellect or imagination: the fifth factor. It has traits such as philosophicalness, complexity, and creativity versus shallowness and conventionality. This factor also has another name, openness to experience.

The Big Five model has been researched and validated by many different psychologists and are at the core of many personality questionnaires. According to

McCrae & Costa ([14]), the extraversion factor can be further refined into six facets: warmth, gregariousness, assertiveness, activity, excitement seeking, and positive emotions.

In this study, we applied the big-five personality model in gaming and focus on the largest trait: extraversion.

3 Preliminary Observations

An online survey was conducted to investigate the relationships between personality traits and enjoyment of computer game play. The survey questionnaire contains three types of questions: questions about player's demographics and gaming experience, questions about personality traits, and questions about enjoyment of playing a particular computer game.

Game enjoyment was measured by an 11-item instrument proposed by Fang et al. ([18]).

Personality traits were measured using the 50 item IPIP ([19]) inventory available at http://ipip.ori.org/. Since its inception in 1999, IPIP has been used in over 60 studies and translated in over 20 languages.

In the beginning of the survey, a participant was first asked to answer questions about his/her demographics and gaming experience such as how often and how long he/she has played computer games. Then the participant would answer 50 questions about his/her personality. Upon finishing the personality questions, the participant was instructed to choose one or more games that he/she regularly plays and assess his/her play experience by answering questions about game enjoyment. If the participant chose more than one game, only one game could be assessed at a time. All the personality and enjoyment questions were randomized for each participant to avoid order effect. Participants were not allowed to skip any questions.

Table 1. Demographic Information of Participants

Variables		
Gender	Male (%)	85.2
	Female (%)	14.8
Culture	American (%)	85.6
	Korean (%)	10.3
	Chinese (%)	4.1
Age	Mean	25.7
	Std.	6.99
How long have you been playing computer/video games?	Mean (years)	14.8
	Std.	8.94
How many hours on average do you play?	Mean	2.81
	Std.	2.126
How often do you play computer/video games?	Daily (%)	41.3
	Weekly (%)	40.7
	Monthly (%)	9.0
	Seldom (%)	9.0

Table 2. Pair-Wise Comparisons of Extraversion Scores Among Different Game Titles

Genre	N	Subset		
	1	1	2	3
SidMeiersCivilization	15	2.566667		
Sims	23	2.800000	2.800000	
Rainbow	13	2.969231	2.969231	2.969231
ResidentEvil	22	3.009091	3.009091	3.009091
NeedforSpeed	33	3.033333	3.033333	3.033333
SuperSmashBros	29	3.041379	3.041379	3.041379
Civilization	14	3.064286	3.064286	3.064286
ElderScrollsIVOblivion	29	3.089655	3.089655	3.089655
WorldofWarcraft	97		3.120619	3.120619
Diablo	44		3.134091	3.134091
Fallout	20		3.140000	3.140000
Quake	14		3.142857	3.142857
DevilMayCry	22		3.159091	3.159091
Crysis	11		3.172727	3.172727
LegendofZeldaTwilight	20		3.175000	3.175000
GrandTheftAuto	54		3.196296	3.196296
WarcraftIII	44		3.204545	3.204545
Halo	66		3.204545	3.204545
CounterStrike	30		3.243333	3.243333
SuperMario	34		3.244118	3.244118
Bioshock	23		3.260870	3.260870
Battlefield	28		3.264286	3.264286
Doom	23		3.265217	3.265217
FIFA	40		3.275000	3.275000
StarCraft	62		3.277419	3.277419
GuitarHero	24		3.291667	3.291667
HalfLife	41		3.302439	3.302439
AssasinsCreed	40		3.310000	3.310000
MarioKart	15		3.313333	3.313333
CommandandConquer	25		3.360000	3.360000
LegoStarWars	14			3.385714
CallofDuty	115			3.396522
MaddenNFL	38			3.397368
FinalFantasy	28			3.425000
SimCity	13			3.438462

Note: The score was out of 5 (maximal score) and p value < 0.05

The survey was conducted in four universities in three different countries: US, Korea, and China. In total, 1096 computer game players responded to the survey. Table 1 presents the descriptive statistics of participants' demographic information.

In order to detect personality differences among players of different games, pair-wise comparisons of extraversion scores among different game titles were performed. The following procedure was used in this analysis: 1) All responses were grouped by game titles. Different editions of the same game title were assigned the same title with the assumption that these different editions should have the similar characteristics and can be categorized as the same kind of game. For example, "Call of Duty", "Call of Duty 2", and "Call of Duty 3" were assigned the same title "CallofDuty". 2) Game titles that were assessed by at least 10 different game players were selected for this analysis. 3) Pair-wise comparisons of extraversion scores were performed among all game titles with at least 10 responses. Table 2 presents the pair-wise comparison results.

Table 2 clearly shows that players of different game titles have different extraversion scores. It indicates that the personality trait, extraversion, may actually affect players' choice of games. To understand the true differences among different game titles, the game falling only in cluster 1 ("SidMeiersCivilization") and the ones falling only in cluster 3 ("LegoStarWar", "CallofDuty", "MaddenNFL", "FinalFantasy", and "SimCity") were compared.

In the cluster 1, "Sid Meier's Civilization" is a turn-based single-player strategy game. It does not involve much social interactions with other characters or players. As shown in Table 2, players of this game title had lower extraversion scores than those of other game titles.

All the game titles falling only in cluster 3 ("LegoStarWar", "CallofDuty", "MaddenNFL", "FinalFantasy", and "SimCity") involve working with a party of people, a team, or other characters. Social interactions with other characters/players were inevitable in these games. Players of all these game titles had higher extraversion scores than those of other games.

A correlation analysis reveals that no significant correlations were found between the extraversion scores and scores of enjoyment-related constructs: affect, behavior, and cognition. The correlation analysis suggests that the differences presented in Table 2 were not related or confound to enjoyment of game players.

The above observations suggest that the personality trait, extraversion, may affect a game player's choice of games requiring different levels of social interactions. In Section 4, we form two propositions about extraversion and computer game play based on these observations and prior research findings.

4 Theoretical Framework

Lazzaro ([17]) identifies four fun keys in a research conducted at XEODesign: 1) Fiero (emotion of triumph over adversity) from the hard fun of challenge and mastery. Players focusing on the game's challenge, strategic thinking, and problem solving favor this fun key. 2) Curiosity from the easy fun of exploration and role-play. Players enjoy intrigue and curiosity. 3) Relaxation from serious fun. Many players enjoy the visceral, behavior, and cognitive experience that games create. 4) Amusement from people fun. Players also use games as mechanisms for social experiences.

One of the four player types proposed by Bartle ([11]) is socialisers who like talking, being part of a group, and helping others.

Based on media enjoyment theories, personality theories, and the technology acceptance model, Fang, Chan, and Nair ([5]) propose a conceptual model of computer game play as depicted in Figure 1.

The prior research on computer game play strongly suggests that social experiences are an important element in the play experience and personality traits may impact on play experience. In conjunction with the observations made in the online survey we conducted earlier, we hypothesize that extraversion, as the largest personality factor in the Big-Five personality model that relates to social interactions, will affect players' choice of games.

Proposition 1: Computer game players who have a high score of extraversion will likely play a game that requires extensive social interactions.

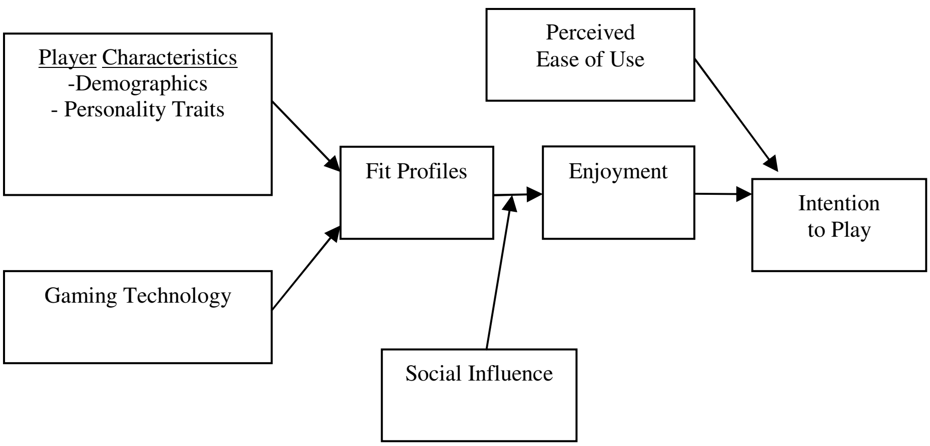


Fig. 1. A framework of computer game play (Fang, Chan, & Nair ([5]))

Proposition. 2. Computer game players who have a low score of extraversion will likely play a game that does not require or requires minimal social interactions.

Based on a series of field studies, Csikszentmihalyi ([20]) creates the flow theory, an understanding of the process by which certain behaviors make life more enjoyable. One element of flow state is the balance between challenge and skill. When user skills are properly matched with the challenges presented in a task, a flow experience will be possible. In computer game play, if we consider a player's extraversion as one skill to interact with other players/characters, a match between the challenge (need for extraversion) presented in a game and his/her skill (extraversion) will likely lead to flow experience and consequently enjoyment. Therefore, game players will likely choose games that requires extraversion compatible to their own personality.

5 Next Steps

To further confirm our observations in an early study and explore the interactions between extraversion and game preferences, we plan to conduct semi-structured interviews to collect qualitative data from game players. Five games are deliberately chosen based on their extraversion scores. These games fall into two different groups based on the early online survey. Two of them are from cluster one (less social: "Sid Meriers Civilization", "Sims 3"), and the rest are from cluster 3(more social: "Call of Dutyz: Modern Warfare2", "Madden NFL 10", "SimCity"). In the interviews, we will investigate how the players would categorize these games based on a game's "personality type" and why.

If the relationships between extraversion and computer game play are confirmed by the interviews, more formal hypotheses will be formed and tested in a controlled environment.

References

1. Entertainment Software Association: The essential facts about the computer and video game industry (2010), `http://www.theesa.com/facts/index.asp` (retrieved December 2010)
2. Griffiths, M.D.: The therapeutic use of videogames in childhood and adolescence. Clinical Child Psychology and Psychiatry 8, 547–554 (2003)
3. Robillard, G., Bouchard, S., Fournier, T., Renaud, P.: Anxiety and presence during VR immersion: A comparative study of the reactions of phobic and non-phobic participants in therapeutic virtual environments derived from computer games. Cyber. Psychology & Behavior 6, 467–476 (2003)
4. Bateman, C., Boon, R.: 21st Century game design. Charles River Media, Inc., Hingham (2006)
5. Fang, X., Chan, S., Nair, C.: An Online Survey System on Computer Game Enjoyment and Personality. In: Proceedings of HCI International (2009)
6. Anderson, C.A., Dill, K.E.: Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. Journal of Personality and Social Psychology 78(4), 772–790 (2000)
7. Fetchenhauer, D., Huang, X.: Justice sensitivity and distributive decisions in experimental games. Personality and Individual Differences 36, 1015–1029 (2004)
8. Douse, N.A., McManus, I.C.: The personality of fantasy game players. British Journal of Psychology 84, 505–509 (1993)
9. Whang, L.S., Chang, C.: Lifestyles of virtual world residents: living in the online game. lineage. Cyberpsychology & Behavior 7(5), 592–600 (2004)
10. Briggs, S.R.: The optimal level of measurement for personality constructs. In: Buss, D.M., Cantor, N. (eds.) Personality Psychology: Recent Trends and Emerging Directions, pp. 246–260. Springer, Heidelberg (1989)

11. Bartle, R.A.: Understanding the limits of theory. In: Bateman (ed.), Beyond game design: nine steps toward creating better videogames. Course Technology, Boston, MA (2009)
12. Fang, X., Zhao, F.: Sensation Seeking, Self Forgetfulness, and Computer Game Enjoyment. In: Proceedings of HCI International (2009)
13. Digman, J., Takemoto-Chock, N.: Factors In The Natural Language Of Personality: Re-Analysis, Comparison, And Interpretation Of Six Major Studies. Multivariate Behavioral Research 16(2), 149 (1981)
14. McCrae, R., Costa, P.: Comparison of EPI and Psychoticism scales with measures of the five factor model of personality. Personality and Individual Differences 6, 587–597 (1985)
15. Peabody, D., Goldberg, L.: Some determinants of factor structures from personalitytrait descriptors. Journal of Personality and Social Psychology 57(3), 552–567 (1989)
16. Tupes, E.C., Christal, R.E.: Recurrent personality factors based on trait ratings. USAF ASD Tech. Rep. No. 61-97, Lackland Airforce Base, TX: U. S. Air Force (1961)
17. Lazzaro, N.: Understanding emotions. In: Bateman (ed.) Beyond game design: nine steps toward creating better videogames. Course Technology, Boston (2009)
18. Fang, X., Chan, S., Brzezinski, J., Nair, C.: Measuring Enjoyment of Computer Game Play. In: Proceedings of the Fourteenth Americas Conference on Information Systems, AMCIS 2008 (2008) ISBN: 978-0-615-23693-3)
19. Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., Gough, H.C.: The International Personality Item Pool and the future of public-domain personality measures. Journal of Research in Personality 40, 84–96 (2006)
20. Csikszentmihalyi, M.: The Evolving Self: A Psychology for the Third Millennium. Harper Collins, New York (1993)

User Modeling Approaches towards Adaptation of Users' Roles to Improve Group Interaction in Collaborative 3D Games

Johanna Renny Octavia, Anastasiia Beznosyk, Karin Coninx, Peter Quax, and Kris Luyten

Hasselt University – tUL - IBBT, Expertise Centre for Digital Media,
Wetenschapspark 2, 3590 Diepenbeek, Belgium
{johanna.octavia,anastasiia.beznosyk,karin.coninx,
peter.quax,kris.luyten}@uhasselt.be

Abstract. This paper focuses on how adaptation of users' roles based on a collaborative user model can improve group interaction in collaborative 3D games. We aim to provide adaptation for users based on their individual performance and preferences while collaborating in a 3D puzzle game. Four different user modeling approaches are considered to build collaborative user models. Through an experiment, we present the validation of these approaches for two different cases: co-located collaboration and remote collaboration. From the experiment, we learned that the *Minimum total time* approach, which defines the best collaboration as the one that gives the shortest total time in completing the task, works mostly effective in both situations.

Keywords: User modeling, adaptation, collaborative 3D games.

1 Introduction

Collaborative 3D games, where players are immersed into a 3D virtual world and can interact with 3D objects or other players, has gained high popularity due to the increasing interest in virtual communities[1] and Massively Multiplayer Online Role Playing Games (MMORPGs)[2]. However, interacting in such 3D world is not always easy for every user. It is more complex compared to real life situations due to the amount of different 3D interaction techniques and interactive tasks that may not always be intuitive for users. Moreover, usage of special 3D input devices with a high degree of freedom (e.g. 3D SpaceMouse) to navigate and manipulate in 3D environments, can cause extra difficulties for the users. This may hinder user interaction and in the end also influence the group collaboration.

We show the user experience in collaborative 3D games can be significantly enhanced by adapting the means of interaction based on the players' model. A model describes the characteristics as well as the capabilities for an individual player. Notice

[1] http://www.secondlife.com
[2] http://www.worldofwarcraft.com

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 668–677, 2011.

the former (characteristics) is more static data, while the latter (capabilities) can evolve over time. The differences between players' models will influence the way the players can collaborate. For example, a large variation in the skills of each user in a collaborative game may decrease the group motivation to collaborate. We believe group adaptation based on individual user's performance can provide more enjoyable group interaction, thus improving group collaboration.

In this paper, we present a study that explores user modeling approaches to construct a collaborative user model based on individual user's performance. We investigate the utilization of the collaborative user model for providing possible adaptations in a collaborative game to benefit the group interaction. The type of adaptation investigated is the assignment of users' roles based on individual performance with respect to the 3D devices being used. For this purpose, an experiment is conducted where two users have to collaborate on a 3D puzzle game using different input devices. To validate the proposed approaches, two different cases are investigated: co-located collaboration and remote collaboration.

2 Related Work

The range of user characteristics relevant to game playing such as ability level, style, and preference, can greatly vary between players. Providing adaptation in games can be considered as a way to accommodate these player differences, maintain engagement and eventually enhance the gameplay experience. A substantial amount of research has attempted to incorporate adaptation in games such as the modification of difficulty levels [1], enemy's behavior [2], or graphic elements of the game environment itself [3]. These investigations mainly focus on providing adaptation based on the state of a single player. Adaptation in collaborative games should be based on the information acquired from all players in the collaborative gameplay.

El-Nasr et al. [4] have defined a set of collaboration patterns based on investigation of cooperative games. One of the identified cooperative patterns is *shared goals*, which is used to force a group of players to work together to reach the same goal. Task division between players in a collaborative game becomes important to improve the collaboration and win the game at the end. Little research has investigated ways of improving collaboration through the act of dividing actions or roles between players. Assigning specific roles to players based on their individual performance and preference, can be considered as a form of adaptation in collaborative games that has not yet been much investigated.

To provide such adaptation, a user model plays a significant role as it contains factual information about the user (e.g. interaction patterns, preferences, abilities) that can be useful to determine the adaptation. User modeling in games has been overlooked, yet considered to have much potential to result in practical benefits for computer game players [5]. Several researches have proposed ideas towards user modeling in games [6]. User modeling in games can be carried out by measuring a player's game ability to describe how well he/she is playing the game and what influences his/her play. Modeling of players accurately is considered to be a crucial aspect in realizing an effective adaptive game. Therefore, it is suggested for game developers and researchers to more consciously using user modeling approaches to model players in a game design and development.

3 Proposed User Modeling Approaches

Our work proposes to build a collaborative user model that aims to provide adaptation of task division between players in a collaborative 3D game to improve their group interaction. The type of collaborative game investigated in this study is a puzzle game with the shared goals collaboration pattern [4]. We have developed a collaborative 3D puzzle game to validate the user modeling approaches. This type of game involves two main actions: rotation and translation.

To construct the collaborative user model, several user modeling approaches are explored. In this study, the user model is defined as the combination of action and device that is predicted to give the best group interaction. We refer to this combination of action and device as the so-called role. To illustrate how the constructed user model can be, Figure 1 shows two actions (*Rotation* and *Translation*) and two devices (*SpaceMouse* and *Phantom*), which are combined to build the user model. Hence, four possible combinations of action and device, or four roles, can be formed: *Rotation with SpaceMouse*, *Translation with SpaceMouse*, *Rotation with Phantom* and *Translation with Phantom*, as a component for the collaborative user model. The other component of the user model is the information about which player is assigned which one of these roles. Having this information, the collaborative user model is completely constructed.

Depending on how the best group interaction is defined, four approaches of user modeling are proposed:

1. **Minimum total time**; the best collaboration is defined as the one which gives the minimum total time in completing the collaborative task. For every pair, the total time is estimated for all possible combinations of action and device and then the combination with the shortest estimated total time is selected. To estimate the total time, for every user, the time to perform each role separately is calculated. Then for all combinations, the total time to perform both roles collaboratively is calculated.
2. **Exclusion of worst individual performance**; the best collaboration is defined as the one which maximizes the group performance by excluding the worst individual performance of a certain user. For every role, the performance difference between users within the pair is calculated. The maximum difference is used to rule out the worst performance value. In the combination with the worst performance, the user with the lower completion time is assigned with the role that was the worst performance of the other user, who obtained the complementary role.
3. **Minimum performance gap**; the best collaboration is defined as the one which results in the most equal performance among the users by minimizing the performance gap between them. For every possible combination, the difference of task completion time between users within the pair is calculated. The combination with the minimum time difference between users is selected as the best combination, which is considered to give the most balanced performance.

4. **Maximum preference**; the best collaboration is defined as the one which makes the best use of users' preference by assigning the most preferred role to each user. For every possible combination, the total subjective preference rating given by users is calculated. Based on the total ratings, the combination with the maximum value is determined as the best combination.

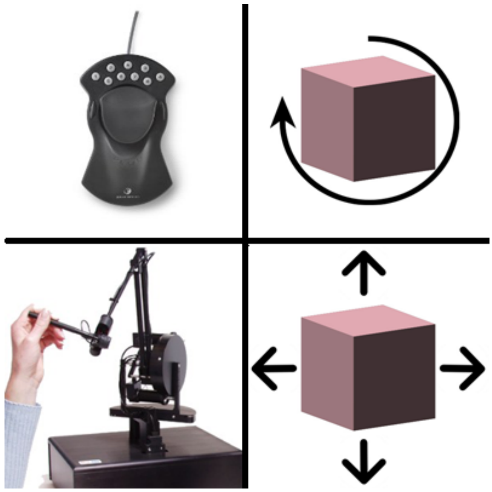


Fig. 1. Actions and devices combined to form the collaborative user model (clockwise from upper right): Rotation, Translation, Phantom, SpaceMouse

In all approaches, for the best combination found, the total task completion time is estimated as follows. First, for every user, the time to perform each role separately is calculated. Then, the total time to perform both roles collaboratively is calculated. The estimated total times are used to evaluate the efficacy of these four user modeling approaches, which will be validated through an experiment described in the next section. With the experiment, we would also like to investigate how the assignment of users' roles can be adapted for two situations: co-located collaboration and remote collaboration. In the co-located case, we analyze a situation where both devices are available for the users so they can switch devices depending on which device are found to be most suitable for them to operate. However, the collaboration often takes place over distance, so switching devices becomes impossible. Therefore, with the second case, we would like to investigate the situation where the availability of a certain device becomes limited to users due to the remote setting. Based on the findings from these cases, the user modeling approaches are compared to determine which approach suits best for both cases.

4 Experiment

Our work investigates the possibility of providing adaptation within a collaborative virtual environment based on a collaborative user model, which assigns users' roles based on individual performance with respect to the available 3D devices. An experiment was conducted to validate the user modeling approaches proposed for constructing the collaborative user model. Several collaborative user models are developed, based on the goal of collaboration, and applied to two different situations: first, all involved devices are available for both users, and second, only one of the devices is available for each user.

The experiment is based on our previous study [7], which showed no interaction effect between different 3D input devices within a heterogeneous setup when freely collaborating in a virtual environment. In this paper, we explore how different users can align their actions while collaborating using heterogeneous setups. Our experiment limits the type of actions a user can perform in a 3D environment to force them to collaborate for reaching a predefined goal. We use the term role-based collaboration to indicate a user can only perform actions related to the assigned role. We believe that applying the role-based collaboration, which explicitly separates roles between users based on the devices, will improve the group performance compared to the free collaboration, where no roles are explicitly assigned.

4.1 Hypotheses

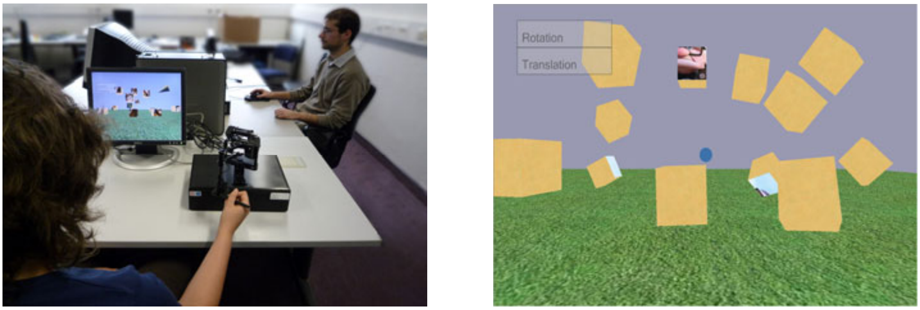
To validate the proposed collaborative user modeling approaches, two hypotheses were suggested: (H1) The modeled task completion time will not differ from the actual[3] time; and (H2) The actual task completion time will be lower than the time during the free collaboration.

4.2 Methods

Twenty unpaid volunteers (16 males and 4 females) were recruited as participants and randomly coupled in pairs for the experiment. The average age of participants was 28 years old, varying from 23 to 34 years old. All participants were people with a computer science background and had little experience working with the devices. All of them were right-handed and used their dominant hand to operate the devices.

The setup described in [7] was used for the experiment. As output devices, two 19" monitors were used. Phantom and SpaceMouse were used as input devices. For both cases, we used the same setup where participants were located in the same room as shown in Figure 2(a), thus co-located. However, participants were seated in such a way that they were not able to see their partner's screen, to simulate a remote setup.

[3] As a result of role-based collaboration.



(a) (b)

Fig. 2. Experiment setting: (a) The experiment setup (b) The experimental task

Participants were asked to collaborate on a 3D puzzle solving task with the shared goal of assembling a complete picture. Each puzzle consists of 12 cubes dispersed in a virtual environment as shown in Figure 2(b). One cube was already placed and served as a visual cue. Part of the picture was presented on one of the cube's sides. Participants were represented by cones with different colors in the virtual environment. Two devices, SpaceMouse and Phantom, were used. Two roles, Rotation and Translation, were defined in the experiment.

The experiment consisted of three independent parts resulting in ten puzzles to be solved. The first part of the experiment corresponded to the session of measuring individual performances. The individual performance data of both participants were gathered and used to build a collaborative user model. Every participant had to complete four separate puzzles individually, which included all possible combinations of roles and devices (see Figure 1). The second part of the experiment consisted of two free collaboration sessions, where no role division was involved so participants were able to rotate and translate as well. The last part of the experiment contained four puzzles to be solved collaboratively, but with applying the division of roles to participants. In this part, roles were assigned complementarily (one participant could only translate, the other rotate).

Throughout the experiment, task completion times were measured and participants were allowed to communicate. After each part of the experiment, participants were asked to fill in a questionnaire about their experience and preference. It took approximately one hour for each pair to complete the whole experiment.

5 Results

In this paper, we focus on improving group interaction in a collaborative 3D puzzle game through the adaptation of users' roles based on a collaborative user model. In Section 3, four approaches have been described to model users' performance in a collaborative setup based on different purposes of collaboration. Using two case studies, we investigate how the proposed approaches can be applied and validated by confirming the aforementioned hypotheses. For this purpose, we analyzed the data using *paired-samples t-tests.* The findings are presented separately for two case studies as follows.

5.1 Case Study 1: Co-located Collaboration

In this case study, we investigate the application of the four user modeling approaches in a situation where both devices are available for the users. The two hypotheses formulated in Section 4.1 will be analyzed to validate each approach. Table 1 summarizes the results of the hypotheses analyzed in this first case study.

Effectiveness of collaborative user models. With the first hypothesis, we investigated whether or not the collaborative user models, which were constructed based on the four user modeling approaches, approximate the real performance in a role-based collaboration. Based on this hypothesis, we expect that the estimated task completion times based on the collaborative user models will not differ from the actual times measured from the role-based collaboration.

For every approach, we found that task completion times do not significantly differ across the modeled and actual times, which means that the modeled task completion time does not differ from the actual time measured as a result of role-based collaboration. From this, it can be concluded that **our collaborative user model is a good approximation to the actual performance**. In another way, we can say that all four user modeling approaches proposed in this study can be considered valid to construct collaborative user models.

Effectiveness of role-based collaboration. The second hypothesis demonstrates the potency of role-based collaboration that adaptation of users' roles improves the group interaction and performance. Based on this hypothesis, we expect that the actual task completion times measured during the role-based collaboration will be lower than the times during the free collaboration.

Only the *Minimum total time* approach confirmed that the role-based collaboration to be effective. We found a significant difference between the actual task completion times and the times measured in the free collaboration session. We also observed that the average actual performance times when roles were assigned was *lower* than the average times during the free collaboration session. This indicates that participants spent significantly less time to complete the task when roles were assigned, which demonstrates that **assigning roles to users improves the group interaction and performance**. In conclusion, we confirm the effectiveness of role-based collaboration when employing the first user modeling approach (i.e. identifying the combination which gives the minimum total time in completing the collaborative task).

For the other three approaches, we found that task completion times do not differ significantly across the actual and free times. However, the average actual task completion times was found to be *lower* than the one of the free collaboration. These findings show that the role-based collaboration is not quite effective when employing these user modeling approaches (i.e. *Exclusion of worst individual performance*, *Minimum performance gap*, *Maximum preference*) as it shows **no significant improvement in the group performance**.

5.2 Case Study 2: Remote Collaboration

For the second case study, the same four user modeling approaches are applied. The only difference is since we limit every participant to only have one certain device available, we will have two best predicted combination of action and device for every pair (e.g. one best combination determined when the first participant has the SpaceMouse, and another best combination determined when the first participant has the Phantom). Table 2 summarizes the results of the hypotheses tested in this case.

Effectiveness of collaborative user models. Every approach showed its effectiveness in constructing collaborative user models since we found no significant difference across the modeled task completion times and the actual times measured as a result of role-based collaboration. This indicates that **the collaborative user model constructed has proven to be a good approximation to the actual performance.** Therefore, we can draw the same conclusion as in the first case study, that all four proposed user modeling approaches are valid to construct collaborative user models.

Effectiveness of role-based collaboration. In two approaches, *Minimum total time* and *Maximum preference*, role-based collaboration showed its effectiveness in enhancing the group interaction and performance. This is shown by the significant decrease of completion times, which confirms that the assignment of roles to users can greatly improve the group interaction and performance. Hence, we can conclude that both user modeling approaches (i.e. determining the combination that gives the minimum total time and maximizing users' preference) are effective.

The other two approaches, *Exclusion of worst individual performance* and *Minimum performance gap*, showed that the average actual task completion times was *lower* than the one of the free collaboration but the difference was not significant. In conclusion, role-based collaboration is not quite effective in these approaches since **no significant improvement in the group performance** is observed.

5.3 Comparison of Proposed User Modeling Approaches

Due to the different aims of collaboration, it is obvious that no single best approach will work for every pair of collaborators. However, we are interested to outline which one of the proposed approaches can be mostly appropriate and effective in both situations of collaboration: co-located and remote. We would also like to confirm whether or not the four user modeling approaches can be applied in both situations.

Two-way repeated measures ANOVA showed a significant main effect of the *user modeling approach* on task completion time ($F_{3,57} = 6.17$, $p<0.005$). This indicates a significant difference among the four approaches across both co-located and remote collaboration. Post hoc tests revealed that the average task completion times of the *Minimum total time* approach was significantly lower than of the other approaches ($p<0.05$). We also found that there was no significant interaction effect between the *user modeling approach* and the *type of collaboration*. This finding suggests that all four user modeling approaches can be used in the same manner, no matter in which situation users are collaborating, either co-located or remotely-located.

Table 1. Case study 1: Co-located collaboration

Approach	Statistics	Hypothesis confirmed?
H1: The modeled task completion time will not differ from the actual time		
Minimum total time	t(9)=0.212, p=0.837	Yes
Exclusion of worst performance	t(9)=1.976, p=0.080	Yes
Minimum performance gap	t(9)=0.159, p=0.877	Yes
Maximum preference	t(9)=0.981, p=0.352	Yes
H2: The actual task completion time will be lower than the time during the free collaboration		
Minimum total time	t(9)=2.302, p=0.047 M_{actual} (M=219.3 s) < M_{free} (M=283.5 s)	Yes
Exclusion of worst performance	t(9)=0.465, p=0.653 M_{actual} (M=237.5 s) < M_{free} (M=250.9 s)	No
Minimum performance gap	t(9)=1.291, p=0.229 M_{actual} (M=225.8 s) < M_{free} (M=263.7 s)	No
Maximum preference	t(9)=1.517, p=0.164 M_{actual} (M=219.3 s) < M_{free} (M=267.9 s)	No

Table 2. Case study 2: Remote collaboration

Approach	Statistics	Hypothesis confirmed?
H1: The modeled task completion time will not differ from the actual time		
Minimum total time	t(9)=1.424, p=0.171	Yes
Exclusion of worst performance	t(9)=1.747, p=0.097	Yes
Minimum performance gap	t(9)=0.996, p=0.332	Yes
Maximum preference	t(9)=1.417, p=0.173	Yes
H2: The actual task completion time will be lower than the time during the free collaboration		
Minimum total time	t(9)=2.776, p=0.012 M_{actual} (M=215.3 s) < M_{free} (M=270.0 s)	Yes
Exclusion of worst performance	t(9)=1.635, p=0.119 M_{actual} (M=237.0 s) < M_{free} (M=270.0 s)	No
Minimum performance gap	t(9)=1.866, p=0.078 M_{actual} (M=233.9 s) < M_{free} (M=270.0 s)	No
Maximum preference	t(9)=2.579, p=0.018 M_{actual} (M=219.4 s) < M_{free} (M=270.0 s)	Yes

We can conclude that, all in all, the *Minimum total time* is the most effective and appropriate approach to be employed in both situations. However as previously mentioned, no single approach will work best since every group may have different goals of collaboration. Therefore, other models can be also widely applied based on the goal of collaboration. They may not guarantee the best performance time but will take into account other important aspects of collaboration (e.g. preference, equal performance, etc.). Although the user modeling approaches have only been validated in a 3D puzzle game, we believe that these approaches can be applied to a wider range of collaborative 3D games.

6 Conclusion

We have presented an investigation of adaptation of users' roles based on the availability of devices to enhance group interaction in a collaborative 3D game. Four different approaches to build collaborative user models were proposed and validated through an experiment. These models were based on different purposes of collaboration. We presented a detailed analysis of every approach for two situations: co-located collaboration and remote collaboration. The *Minimum total time*, a user modeling approach by determining the combination that gives the minimum total time, is found to be the most effective approach in both situations. We have demonstrated that incorporating adaptation of assigning roles to users based on a collaborative user model built using this approach, improves collaboration between two users.

References

1. Chanel, G., Rebetez, C., Bétrancourt, M., Pun, T.: Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In: Proc. MindTrek, pp. 13–17 (2008)
2. Merrick, K.: Modeling motivation for adaptive nonplayer characters in dynamic computer game worlds. Comput. Entertain. 5(4), 1–32 (2007)
3. Buttussi, F., Chittaro, L., Ranon, R., Verona, A.: Adaptation of graphics and gameplay in fitness games by exploiting motion and physiological sensors. In: Proc. SG, pp. 85–96 (2007)
4. El-Nasr, M.S., Aghabeigi, B., Milam, D., Erfani, M., Lameman, B., Maygoli, H., Mah, S.: Understanding and evaluating cooperative games. In: Proc. CHI (2010)
5. Storey, D.: Modelling user ability in computer games. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS (LNAI), vol. 3538, pp. 521–523. Springer, Heidelberg (2005)
6. Charles, D., McNeill, M., McAlister, M., Black, M., Moore, A., Stringer, K., Kücklich, J., Kerr, A.: Player-centered game design: Player modelling and adaptive digital games. In: Proc. DiGRA, pp. 285–298 (2005)
7. Beznosyk, A., Raymaekers, C., Coninx, K., Quax, P., Lamotte, W.: Investigating the influence of communication and input devices on collaboration in virtual environments. In: Proc. GRAPP (2010)

MusicTagger: Exploiting User Generated Game Data for Music Recommendation

Hannes Olivier, Marc Waselewsky, and Niels Pinkwart

Clausthal University of Technology
Tannenhöhe 25, 38678 Clausthal-Zellerfeld, Germany
{hannes.olivier,marc.waselewsky,niels.pinkwart}@tu-clausthal.de

Abstract. The system "MusicTagger" is a game in which two players hear 30 seconds of a song, describe it independently and get points if they succeed in making the same descriptions. Additionally, it is a music recommendation system which compares songs with the help of the descriptions given in the game. MusicTagger is based on the principle of "human computation", meaning that problems (in this case, music recommendation) are solved by computers via eliciting human knowledge and making intelligent use of the aggregated information. This paper presents the design and implementation of the "MusicTagger" system together with results of an empirical lab study which demonstrates the potential of the recommendation engine.

1 Introduction

Deciding what music people would (potentially) like or not is a challenging but potentially highly valuable task for a computer system. There is quite a variety of attempts for creating good music recommendation systems. A prominent example is Apple Genius which is able to generate playlists of music similar to a selected song, relying essentially on collaborative filtering. Other approaches for music recommendation include the process proposed by Eck et al. [1], which relies on four steps for calculating song similarity: first, an acoustic feature extraction is calculated. After that, tags to the songs will be categorized. Next, AdaBoost, a meta-learning algorithm, is used for tag prediction. While not outperforming social tagging, first tests indicate a decent success of this approach [1]. Logan [2] used acoustic information for calculating a distance between songs. When new songs are integrated into a song collection, they will first be grouped to a song set which has the minimal distance. Yoshii, Goto, Komatani, Ogata and Okuno [3] used a hybrid music recommendation method. They tried to reduce the weaknesses of content-based recommendation and collaborative filtering by combining both approaches. Kuo, Chiang, Shan and Lee [4] designed an emotion-based music recommendation system. Based on their analysis of film music and the emotions this music conveys, they analyzed different music to assign emotional classifications and make recommendations based on this information. The system of presented in [5] is separated into seven segments: a track selector, a feature extractor, a classifier, a profile manager, a recommendation module, an interface and a database. New songs will first be integrated to the first two segments. The track selector categorizes the

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 678–687, 2011.

songs as monophonic music objects or polyphonic music objects. The feature extractor then detects technical data of the songs (e.g., pitch density, pitch entropy, tempo degree and loudness). This data is needed for classifying the songs into music groups. The profile manager then saves information like the last date a user has looked for a song.

The “Friend of a Friend” (FOAF) and “Rich Site Summary” (RSS) vocabularies are used for the system presented in [6]. Here, music recommendations are generated in four steps: (1) get interests from user’s FOAF profile, (2) detect artists and bands, (3) select related artists from artists encountered in the user’s FOAF profile, and (4) rate results by relevance. Finally, TagATune [7] is a game where teams of two players play 3:30 minutes for a round. In this round, they hear seven songs for 30 seconds each. In these 30 seconds, they describe the song they hear and see the descriptions their partner gave. After each round, they have to evaluate if they have heard the same song or not. For a right evaluation, they get points. If they earn enough points, they can play a bonus round. In this, they hear three songs and they have to say which one has the biggest difference to the other two. If both players say the same, they get points. The descriptions in the first rounds and the evaluation of the (in)different songs are saved for a music recommendation system.

The approach for music recommendation presented in this paper is similar to the ESP Game [8] and to TagATune: it is based on the design principle of human computation where essentially humans solve tasks that are hard to do for computers (but where computers come into play in terms of aggregating and intelligently processing the human-generated data). In contrast to TagATune however, our approach aims at producing *categorized* key words as results (for example, artist or instrument), so that the underlying recommendation engine can rely on pre-classified information and can assign different weights to different types of information about songs.

In the next sections of this paper, the system MusicTagger and a pilot lab study conducted with the system to test the game and the recommendation engine are presented.

2 The MusicTagger System

The system is separated into two parts, the game to collect data and the recommender system using the generated data.

2.1 The Game

In the game, teams of two players hear the same part of a randomly chosen song. This part is 30 seconds long and chosen randomly from the length of the song. This approach was taken in order to account for changes in the song over time for obtaining key words (e.g., an intro might have a different style than a refrain).

While hearing the song snippet, players have to describe it to earn points. Figure 1 shows the main game screen. The players have five categories where they can put in their descriptions: genre, instrument, artist, title and miscellaneous. This is different from other music recommendation systems like TagATune that provide only one category. If both players add the same description word to a category, they get points

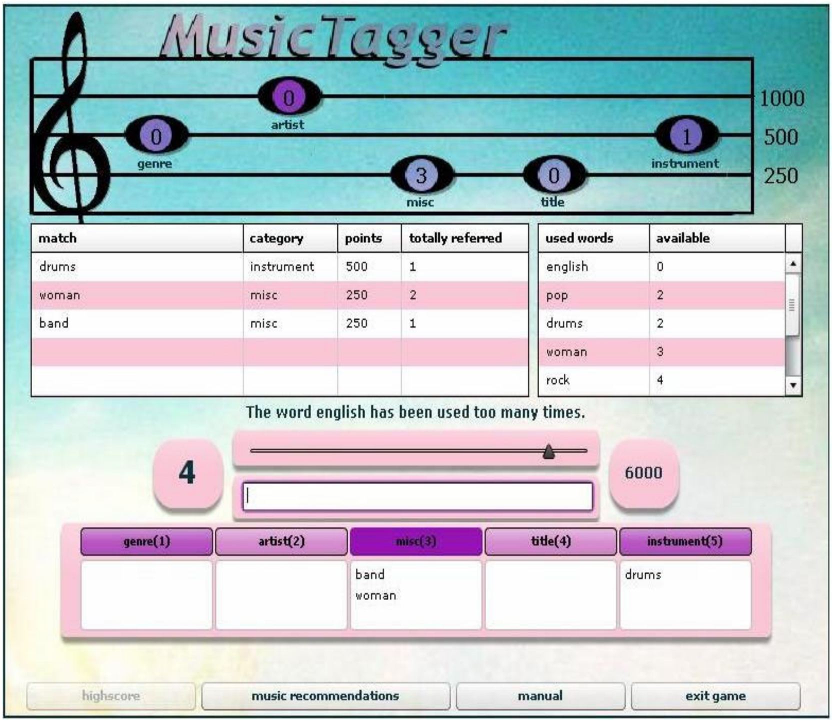


Fig. 1. User interface of MusicTagger game

and can play one more round with each other (and this song description is added to the database for the respective category). Points are accumulated over the different rounds to encourage the players to continue with the game. For each matching word, 1000, 500 or 250 points can be earned. The point assignment is variable and depends on the current state of the database. The category with the fewest data contained for the played song yields the most points (in Figure 1, the category “artist”). This design encourages players to enter data that the system really needs in order to make good music recommendations.

In the user interface, the circles in the top of the screen also contain the number of words the other player already provided for each category (e.g., in Figure 1, one instrument keyword was typed in by the partner). This design was chosen to improve the player’s chances to find a match. A slider represents the remaining time, and the number to the right shows all points earned by the team. Below, the players can provide their descriptions category-wise and they can see which descriptions they have already given in this round (e.g, “drums” for the instrument category).

In a pilot study conducted with a preliminary version of the system, it was noticed that participants gamed the system by describing every song with the same words to earn points, regardless of what the songs were about. This resulted in wrong descriptions for the songs. To discourage this kind of behavior, a “black list” was integrated (table on the right side). The black list counts every word a team has already earned points with. Once a word has been used 5 times by the team, the word is forbidden from then on.

2.2 The Recommendation System

With the data collected by the game, the recommender system compares the songs to each other using the algorithm shown in Figure 2.

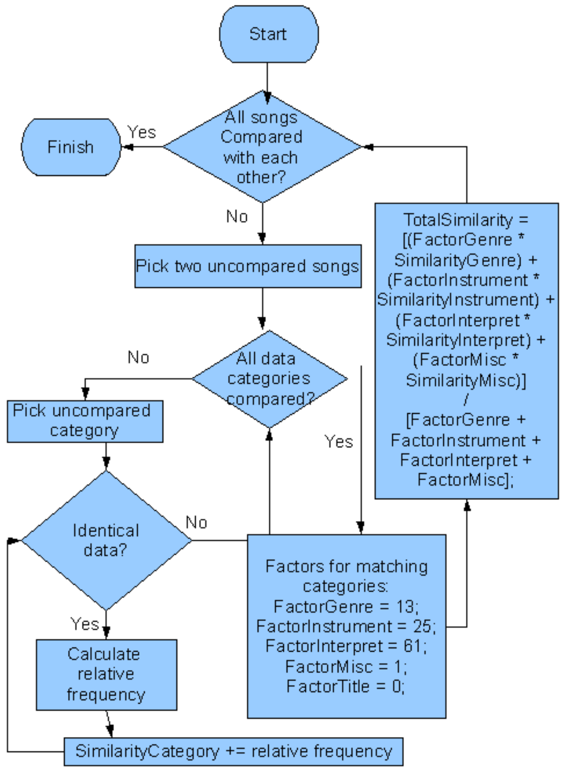


Fig. 2. Similarity calculation algorithm

In the algorithm, songs will be compared if there is at least one category (e.g., “genre”) where both songs have at least one data entry. For each such category, a similarity score will be calculated (which is 0 if the data entries for that category are different, 1 if they are identical, and in between if there are partial matches such as one song having 2 data entries and another having 3, with 1 of them being shared).

An example calculation: Let us assume we have two songs with one having (0, 3, 1, 1, 2) entries by category and the other having (0, 0, 2, 1, 3) entries respectively. Let us further assume that (0, 0, 0, 1, 1) of these entries are identical. In category one, both songs have no data: this has no influence to the total similarity. The second category will not influence the total similarity either because the second song has no entries in this category. Category three has entries for both songs, but no identical description, so its similarity in this category is 0%. The fourth category has only one entry for each song and the entries are identical, so this category has a similarity of 100%. In the fifth category, we first calculate the relative frequency for the identical word: It is 1/2 for the first song and 1/3 for the second song. Next, the smaller frequency is divided by the bigger one, and the result is multiplied with the average of the relative frequencies. The similarity for this category is thus min(1/2 / 1/3, 1/3 / 1/2) * (1/2 + 1/3) / 2 * 100 = 27.78%.

For these (max. 5) category similarities, a weighted average is calculated next. These weight factors are: genre 13, instrument 25, artist 61, miscellaneous 1, title 0 (i.e., the title will not be used to recommend songs, while same artist has a high weight). In our example, the third category (0%) could be the category "artist", the fourth category (27.78% similarity) could be the category "instrument" and the fifth category (100% similarity) could be the category genre. Accordingly, the total song similarity is ((61 * 0) + (13 * 100) + (25 * 27.78)) / (61 + 13 + 25) = 20.15%

Note that the weight of the "song title" information may be subject to debate. Having different versions of the same song in a database might result in the title information being an important category (e.g., in the classical music section, different orchestras may play the same Mozart piece). For calculating the weight factors, 18 song pairs of a small pilot study were used. Participants of this study reported on the similarity of these 18 song pairs manually. The similarities (for each category) for the 18 song pairs and the weight factors for the algorithm were then calculated based on this data.

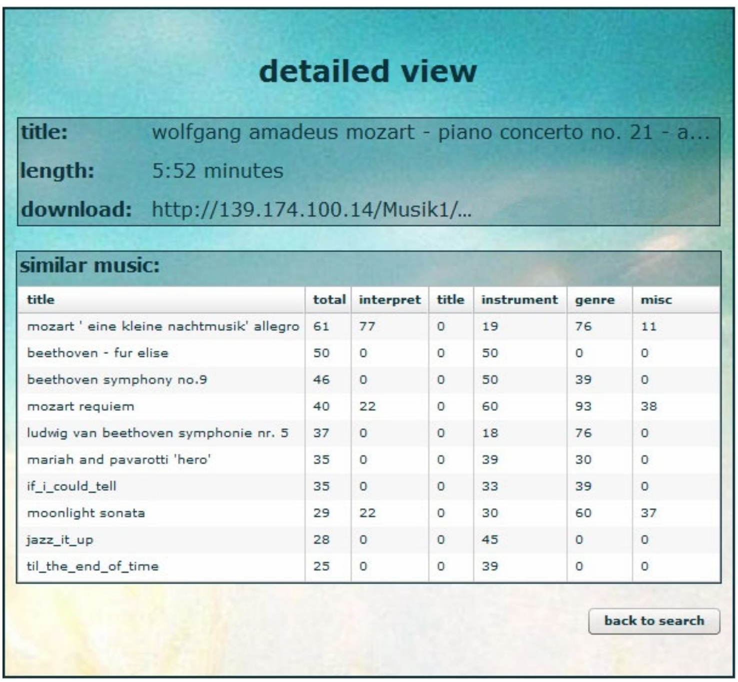


Fig. 3. User interface of music recommendation engine

The user interface of the recommendation system allows users to search for songs in the database. They can perform a free-text search and the system then searches for songs which have the entered name or which have been described in the game with the entered key word. Then, the system shows a result list with relevant songs. By clicking on a result, the system shows the title of the selected song, its length, (optional) a link for downloading the song and a ranking with ten songs which have the highest similarity with the selected song. The table also includes the total similarity and the similarities for each category for all ten results. An example for a song of Mozart is shown in Figure 3. It shows the top-ten-ranking with "Mozart 'eine kleine nachtmusik' allegro" having the biggest similarity (61%) to the chosen song and "til_the_end_of_time" on rank ten with a similarity of 25%.

3 Research Questions

The MusicTagger system was evaluated in an empirical lab study in order to answer several research questions. One of the main problems of the previous pilot study with the preliminary version of the system was people cheating by using certain (fixed) words all the time to continue playing. The new version of the game now included a black list. We were interested whether the black list changed the players' behavior and whether its use resulted in different answers. Also, we were generally interested if the players enjoyed the game.

Concerning the recommendation system, we were interested in the following research questions:

- Do the similarities calculated by the recommendation system correlate with user generated similarities?
- Do the users accept the systems recommendations?
- Does more data generate better recommendations?

4 Study Description

For testing our system, a two-week study with 44 participants was conducted. In the first week, the participants played the game; in the second one they tested the recommendation engine. Not all participants showed up in week two, resulting in only 41 people evaluating the recommendation system. The participants were not informed that the two study sessions they participated in used the same underlying system. Most of them did not see the relation of the two sessions. Some even asked if these were two independent studies or if there was some small connection. The participants were paid 25 Euros. In both weeks, five sessions with 8 to 10 participants were conducted. The sessions in the first week needed an even number of participants, so that every participant had a partner to play with every time. In the second week, the participants worked alone, so the number of participants per session did not matter. All participants were in the same room, but they were sitting separated from each other and were not allowed to communicate with each other. This was enforced through the experimenter who was sitting in the room.

The database was filled with 102 songs. This number was chosen on purpose to be able to collect enough data to evaluate the recommendation system (too few songs will lead to repetitions, too many songs will lead to too few keywords for each song). We used music out of eleven genres (black, jazz, classic, dance, hip hop, metal, pop, rap, rock, alternative and folk music) with 9 songs for each genre. Three songs were duplicated (to check for the system-generated similarity for identical songs).

In the first week, the participants played the game and, as a side-effect, filled the database with information. Two versions of the game were played (one with blacklist, the dynamic category-dependent point system and some UI improvements, one without) to check whether the black list would improve the results of the answers. Everybody played two sessions (one with each version of the software) for 30 minutes. After the two sessions, the participants had to fill out a survey to evaluate both versions.

After this, the database was filled with 1100 descriptions (excluding entries from the “control version” that did not have the black list) for the 102 songs. Only 91 songs got descriptions from the participants, the other 11 did not get any (some of them were not played, some did not lead to scores).

To answer the question if the black list produced better results, the generated data from both versions was compared. The data of the non-blacklist version was also merged with the data from a previous study data to see if more data generates clearer results in the recommender system.

In the second week, the recommendation system was reviewed. 32 new pairs of songs (i.e., not containing the songs that belonged to the initial 18 pairs that were used to inform the algorithm) were prepared, with every song having at least one entry in the categories “genre”, “instrument” and “miscellaneous”. These 32 pairs belonged to four sets: eight pairs with the same artist, eight pairs with different artists and a system-calculated similarity >70%, eight pairs with different artists and a similarity < 30%, and eight pairs with different artists and a similarity between 30% and 70%.

Every participant was given one pair out of every set. The participants first listened to both songs of all pairs and then had to evaluate the similarity of their four pairs (in the categories genre, instrument and total similarity) on a scale from 0% to 100% in steps of 10%. After this, they were presented with the similarity results of the system for their four pairs and had to write down whether they thought the results of the system were satisfying or not. They described this by their own words in a free text. Each of the 32 test pairs was reviewed by at least 5 participants.

5 Results

Concerning the impact of black list, some results were found. 75% of the participants stated that the list influenced their gaming. They said that as they were looking at the “black list”, they tried to generate less common entries and not always standard words like rock, pop, guitar. These general terms were “reserved” for game situations where no other similarity was found or when the category offered a lot of points. The people that stated that they were not influenced by the black list said that they either never had the problem of blocked words or that they didn’t know a lot of words and accepted the risks of an early game end.

The database entries did not show a significant change in the overall use of “standard” keywords. The only observable tendency is that in the “black list” version, these words were used for fewer songs, which is in line with the statements of the participants.

The questionnaire at the end of the first study week contained some questions concerning the differences of the systems. The participants preferred the new, flexible point system. In the control version, the point table was not as flexible as it was in the experimental version and contained static points by category. The participants said that they were curios in each round which category would bring the most points and tried to reach at least one match in this category. This, of course, supports the design choice of making the points dependent on the state of the database (assigning many points to data entries that the system needs to make good recommendations).

Overall, the players enjoyed the game and gave average scores of 2.0 for motivation, 2.1 for usability and 2.1 for fun (on a scale of 1(good) to 5(bad)). These scores were assigned for the features of control system version. Additionally, the participants were asked about their opinion regarding the differences between the control and the experimental version. Here, they stated that they preferred the experimental version.

The similarities reported by the participants in the second week of the study were used to evaluate the recommender system. Comparing the system-generated and the user-generated similarities resulted in interesting insights.

The recommendation algorithm could only compare two of the three songs doubled in the database. The third one didn’t receive descriptions. When the two songs were compared the system generated a similarity of 91 and 92%.

An example for a pair with a good result in the total similarity was a pair with two songs of the artist “Scooter”. The system calculated a total similarity of 65% (100% artist, 0% genre, 37% instrument), the participants evaluated this song pair with 70% (on average) for all categories. An example for a poor result is one pair which consists of two songs from a German artist called Heintje. The system calculated a total similarity from 24%, but the participants stated a total similarity of 76%. In this case, the category artist contained no data, so the categories genre and instrument became very important for calculating the total similarity. But for both categories, one song had only one description entry and the other song had a few different descriptions, so that the category similarities were also low and finally, the total similarity became very low. This shows the problem of small data sets – with more data in the database (particularly with artist information), the calculation would have been more accurate. The current algorithm was able to produce an average difference to user generated evaluations of 15.4 percent (with sd=12.1).

The question if the participants were satisfied or dissatisfied with the similarity results delivered of the system was analyzed based on the free-text answers. In total, the participants answered they were satisfied with a result 106 times (65%), “still satisfied” 12 times (7%) and dissatisfied 46 times (28%). In interpreting this data, it has to be considered that sometimes humans do not agree at all: in one extreme case, one participant stated a pair similarity of 80% while another participant stated 20%. With similarity estimations diverging this far, an automated algorithm cannot satisfy both participants.

When the participants were asked about the top ten recommendations, they stated that the recommendations were overall very good. In most cases, only one or two songs were considered wrongly recommended. Here, one has to consider that in our study, only nine songs per category were available – so it is not surprising that a top ten list contains a few songs that not very similar.

By comparing the difference between the similarity score assigned by the users and the one calculated by the system, we additionally observed the following result:

- If the difference was 10 percentage points or less, the participants were always satisfied with result of the system.
- If the difference was between 11 and 15 percentage points, many, but not all participants were satisfied with the result of the system.
- If the difference was more than 15 percentage points, almost every participant was dissatisfied with the result of the system. Only in a few cases, when the total song similarity was less than 30%, a few participants said there were satisfied, because of the same tendency of low similarity. With higher song similarities, the participants looked more for the numerical value and were dissatisfied, because of the high difference.

According to this (relatively tough, but realistic) measure, we have three classes (<10%, 10-15%, >15%) in terms of difference between system and user-stated similarity. For our 32 song pairs, this classification leads to 14 satisfying results (44%), 8 mostly satisfying results (25%), and 10 bad results (31%), cf. Table 1 for details.

Table 1. Differences between system-calculated and user-stated similarity

	Satisfying (<10%)	Still satisfying (10-15%)	Dissatisfying (>15%)
Set 1: same artist	3	2	3
Set 2: different artist, >70% system similarity	3	2	3
Set 3: different artist, <30% system similarity	6	1	1
Set 4: different artist, 30%-70% similarity	2	3	3

6 Conclusion

In this paper, we report on a pilot study with the system "MusicTagger", a two-part-system consisting of (1) a game which is based on the design principle of human computation, and (2) a music recommendation system which uses the descriptions delivered by game players for calculating music similarities. Users can search for a song they like, and based on their choice the ten songs with the highest similarity to the selected song are recommended.

The study showed that, even after relatively few hours of generating data through playing (and, thus, based on relatively few data), the users were satisfied with the game and with the results of the recommendation engine. Specifically, the dynamic allocation of game points based on the data entry types that the recommendation engine could benefit from and the “black list” which prevents gaming the system were successful. In the recommendation system, the majority of users were satisfied with the music similarity results delivered by the system (yet, of course larger studies with more songs and users will be required to confirm this result).

References

1. Eck, D., Lamere, P., Bertin-Matiuex, T., Green, S.: Automatic Generation of Social Tags for Music Recommendation. University of Montreal
2. Logan, B.: Music Recommendation from Songs Sets. Cambridge Research Laboratory, HP Laboratories Cambridge, HPL-2004-18 (2004)
3. Yoshii, K., Goto, M., Komatani, K., Ogata, T., Okuno, H.: Hybrid Collaborative and Content-based Music Recommendation Using Probabilistic Model with Latent User Preferences. University of Victoria (2006)
4. Kuo, F.-F., Chiang, M.-F., Shan, M.-K., Lee, S.-Y.: Emotion-based Music Recommendation By Association Discovery from Film Music (2005)
5. Chen, H.-C., Chen, A.L.P.: A music recommendation system based on music data grouping and user interests. Department of Computer Science, National Tsing Hua University (2001)
6. Celma, O., Ramírez, M., Herrera, P.: Foafing the music: a music recommendation system based on RSS feeds and user preferences. University of London, Queen Mary (2005)
7. Law, E.L.M., von Ahn, L., Dannenberg, R.B., Crawford, M.: TagATune: A game for music and sound annotation. Austrian Computer Society (2007)
8. von Ahn, L., Dabbish, L.: Labeling Images with a computer game. Vienna, Austria (2004)

The Influence of Social Experience in Online Games

Hua Qin[1], Pei-Luen Patrick Rau[2], and Song-feng Gao[1]

[1] Department of Industrial Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, P.R. China
[2] Department of Industrial Engineering, Tsinghua University, Beijing 100084, P.R. China

Abstract. The objective of this study is to explore the influence of social behaviors in the online games on the behaviors in the real world. Four kinds of social behaviors in the online games such as chatting, making friends, role playing, team work are proposed, which these behaviors tie virtual world and real society. Based on the data from 101 effective respondents, this study analyzes the relationship between average playing time, social behavior in the real world and social behaviors in the game world. The results find that social behaviors in the game environment change significantly players' social behavior in the real world, and also contribute largely to more behavior repetition.

Keywords: Social experience; online games.

1 Introduction

Every day, millions of players interact with other players in online environments known as Massively-Multiplayer Online Role-Playing Games (MMORPG) [1]. Playing computer online games has become a social experience [2]. The shared experience in the virtual society, the collaborative nature of most activities, makes players different experience from in the offline games [3,4]. Despite many researchers focus on the influence of online games such as addiction, flow, lonely and depression and so on [5,6,7], a few researches the relationship between social experience in the online game and in the real world: do social experience in the online game work on playing time? Which social elements in the online game influence real lives of players?

The aim of this study is to explore the influence of social behaviors in the online games on the behaviors in the real world. Next, some related literatures would be introduced.

2 Related Works

Designers of the games want to promote the social natures, as they believe that these encounters are essential to the success of the games. And players' social experience in the games can increase the appeal and longevity of the game [8]. Through an empirical model of player motivations, Yee [9] researched the contribution of social component in the online games to the motivation.

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 688–693, 2011.

Commonly, social experience in the games includes casual chat, helping other players, making friends, collaboration, groups and so on [9, 10]. In terms of the literature and characteristics of the online games, four main components are proposed in this study. The components will now be described briefly:

1) **Chatting**—Communicating with other players through text-chat mode.
   In course of playing games, players often share skills and expertise to accomplish the game's objectives. While asking questions, teaching other players in the game players often use the private "tell" mode of the text-chat interface [11, 12].
2) **Teamwork**—Collaborating with other players to accomplish tasks in the game world and deriving satisfaction from being part of a group effort.
   The players have to accomplish some tasks are purposefully too difficult for a single character, and require the help of a group of other players [13]. If players choose to cooperate with others in the same situation, quest rewards are more attainable and can be accumulated faster [14].
3) **Role-Playing**—Acting as a persona in the background story and interacting with other players to create an improvised story.
   In the games, players take part in the narrative flow of the virtual world and shape it through their participation in the world [15].
4) **Making-friends**—Using the online environment to know other players and keep touch with the persons.
   Researchers found that groups of friends move from one game to another [13].

Commonly online games are often described as extremely time-consuming. The average play time per character over a week was 10.2 hours [16]. Researchers found that impacts of guild membership on playing time significantly [16].

3 Methodology

In order to find the impact of social behaviors in the online games on behaviors in the real world, a survey was conducted.

3.1 Instruments

The survey was consisted of three parts, which were social experience in online games, social behaviors in the real world, and behavior repetition. The instrument measuring social experience in online games includes four dimensions, which are chatting, making friends, role playing, team work. The instrument measuring social behaviors in the real world comprises three sides such as communication, life style, and making friends. Items of the two instruments were measured with seven-point Likert scales between disagrees strongly and agrees strongly.

Through asking participants how many hours they spent averagely every day on the games represented behavior repetition.

3.2 Participants

Young male persons were considered as majority of online game players [17]. Therefore, in this study college male students were recruited as the major. There were 160 male participants who answered the questions. Only 101 data from the respondents were effective. Among these effective respondents, approximate 93% of them like playing online games. And about 45% of them play games more than 3 years.

3.3 Procedure

A survey was conducted in April, 2010. The whole survey included two steps. The first step was to investigate demographic information. Demographic information collected educational level, age, gender, and experience of game. The Second step was participants answering the instruments of social experience in the online games, social behavior and behavior repetition in the real world.

4 Results

In order to examine and explore how degree that chatting, making friends, team work and role playing in the game world affect players' life in the real world, correlation and regression analysis are completed. Independent variables are chatting, making friends, team work and role playing in the game world. Dependent variables are average playing time of each time and social behavior in the real world.

Table 1 shows the correlation between dependent variables and independent variables. In the table, playing time of each day is related significantly with Chatting and Role-playing. And Chatting, Making-friends, Role-playing, Team-work in the game world are impacted evidently on the social behaviors in the real world

Table 1. Correlation

Variables		Chatting	Role	Team	Friend
Time	Pearson	.207	.266	.093	-.025
	Sig.	.037	.007	.357	.801
Real	Pearson	.347	.374	.485	.286
	Sig.	.000	.000	.000	.004

* 'Time' represents average playing time of each day. 'Real' represents social behaviors in the real world.

In order to explore in what extent the social experience in the game world affect social behaviors in the real world and behavior repetition, regression analysis was computed. Table 2-5 present the model summary and coefficients of the regression. For behavior repetition, adjusted-R-square of the model is 0.543. On condition that other variables do not change, the standardized-coefficients imply that playing time increases 0.827 units while Chatting increases 1 unit. In the same way, playing time increases 0.609 units while Role-playing increases 1 unit. Playing time increases 0.471 units while Team-work increases 1 unit. And playing time decreases 0.569 units while Making-friends increase 1 unit. For social behaviors in the real world, adjusted-R-square of the model is 0.896. On condition that other variables do not change, the standardized-coefficients imply that playing time increases 0.468 units while Chatting increases 1 unit. In the same way, playing time increases 0.176 units while Role-playing increases 1 unit. Playing time increases 0.082 units while Team-work increases 1 unit. And playing time increases 0.439 units while Making-friends increase 1 unit.

Table 2. Model Summary-Behavior repetition

Multiple R	R^2	Adjusted R^2
0.894	0.799	0.543

Table 3. Coefficients-Behavior repetition

	Standardized Coefficients		df	F	Sig.
	Beta	Std. Error			
Chatting	0.827	0.073	16	127.276	0.000
Role	0.609	0.075	17	65.265	0.000
Team	0.471	0.071	10	43.569	0.000
Friend	-0.569	0.072	13	61.822	0.000

Table 4. Model Summary-Social behaviors in the real world

Multiple R	R^2	Adjusted R^2
0.951	0.906	0.896

Table 5. Coefficients- Social behaviors in the real world

	Standardized Coefficients		df	F	Sig.
	Beta	Std. Error			
Chatting	0.468	0.053	1	78.205	0.000
Role	0.176	0.043	2	16.418	0.000
Team	0.082	0.032	3	6.384	0.001
Friend	0.439	0.043	1	106.210	0.000

5 Discussion and Conclusion

For players' playing time, Chatting and Role-playing are the main factors. This result indicates that communicating with other players and acting as virtual persons contribute to keeping players playing online games much more. Players have interests in helping and chatting with other players. And also they derive satisfaction from being part of the game world.

For social behavior in the real world, the important factors are Chatting and Making-friends, and next factor is role playing in the virtual world. This implies that form long-term meaningful relationships with others and playing a role in the game change players' real life style. Generally, online games are massive multiplayer playing. In order to fulfill their tasks in the games, players have to constantly seek workarounds and external support [10]. Therefore, social behaviors oriented on tasks in the games are different from in the real world. But these actions in the game influence significantly those in the real world.

Commonly, online games are becoming massive multiplayer playing. Thus, part of the games develops to socialize. In order to completing tasks in the games, players have to seek external support including from game world and real world. Therefore, it is necessary to study the influence from the social behaviors in the game world. This study finds that for players' playing time, Chatting and Role-playing are the main factor. And for social behavior in the real world, the four factors all have contribution. Thereinto, Chatting, Role-playing, and Making-friends play important roles.

References

1. Woodcock, B.S.: MMOG chart, `http://www.mmogchart.com/` (accessed September 21, 2010)
2. King, B., Borland, J.: Dungeons and dreamers: the rise of computer game culture. Graw-Hill, New York (2003)
3. Jakobson, M., Taylor, T.L.: The Sopranos meets EverQuest: social networking in massively multiplayer online games. In: Proceedings of the 2003 Digital Arts and Culture (DAC) conference, Melbourne, Australia, pp. 81–90 (2003)
4. Yee, N.: Ariadne - Understanding MMORPG addiction, `http://www.nickyee.com/hub/addiction/home.html`(accessed September 21, 2010)
5. Qin, H., Rau, P.P., Zhong, H.Q.: Construction of online game addiction based on player experience. In: the 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore,, pp. 32–436 (2007)
6. Chou, T.J., Ting, C.C.: The role of flow experience in Cyber-Game Addiction. Cyber Psychology and Behavior 6(6), 663–675 (2003)
7. Pace, S.: A grounded theory of the flow experiences of Web users. International Journal of Human-Computer Studies 60(3), 327–363 (2004)
8. Ducheneaut, N., Moore, R.J.: The social side of gaming: A study of Interaction patterns in a Massively Multiplayer Online Game. In: CSCW 2004, Chicago, Illinois, USA, November 6–10 (2004)
9. Yee, N.: Motivations for play in Online games. Cyber. psychology & Behavior 9(6), 772–775 (2006)
10. Manninen, T.: Interaction forms and communicative actions in multiplayer games. Game Studies 3(1) (2003)
11. Ducheneaut, N., Moore, R.J.: Gaining more than experience points: Learning social behavior in multiplayer computer games. In: Proceedings of CSCW (2004)
12. King, B., Borland, J.: Dungeons and Dreamers. McGraw-Hill, New York (2003)
13. Jakobson, M., Taylor, T.L.: The Sopranos meets EverQuest: social networking in massively multiplayer online games. In: Proceedings of the 2003 Digital Arts and Culture (DAC) conference, Melbourne, Australia, pp. 81–90 (2003)

14. Sherlock, L.M.: When Social Networking Meets Online Games: The Activity System of Grouping in World of Warcraft. In: SIGDOC 2007, El Paso, Texas, USA, October 22–24 (2007)
15. Seay, A.F.: Project Massive: The Social and Psychological Impact of Online Gaming. Carnegie Mellon University Pittsburgh, Pennsylvania (2006)
16. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J.: "Alone Together?" Exploring the Social Dynamics of Massively Multiplayer Online Games. In: Montreal, Quebec (2006)
17. Griffiths, M.D., Davies, M.N.O., Chappell, D.: Breaking the stereotype: the case of online gaming. CyberPsychology and Behavior 6(1), 81–91 (2003)

Head-Pose Recognition for a Game System Based on Nose's Relative Position

Qingjie Zhao*, Xiaoming Shi, and Yuxia Wang

Beijing Laboratory of Intelligent Information Technology,
School of Computer Science and Technology, Beijing Institute of Technology
Beijing 100081, China
zhaoqj@bit.edu.cn

Abstract. This paper proposes a head pose identification method using nose's relative position information in a face region and develops a game system based on visual head pose identification techniques to control a virtual robot walking in a virtual maze environment. Adaboost classifiers are used to detect a face and give a rough classification. A nose classifier is used to detect noses in a detected face region. The nose's relative position information is utilized to further determine a specific head pose. A method for selecting the best nose area is also proposed to solve the problem that multi-nose areas may be detected in one face region. Experiment results show that by using these techniques correct head poses can be acquired and applied in the game system.

Keywords: Human computer interaction, head pose, nose position.

1 Introduction

Head pose recognition or estimation is the process of inferring the orientation of a human head from a digital image. The authors in [1] give a survey of head pose estimation methods that have been published from 1995 to 2008, which describes eight categories of head pose estimation approaches and their advantages and disadvantages. These categories include appearance template methods, detector array methods, nonlinear regression methods, manifold embedding methods, flexible models, geometric methods, tracking methods and hybrid methods. The authors in [2] regroup the different works into three classes regarding their approaches: using statistical/classification methods, using geometrical methods and using depth information. A survey is provided in [3] on gesture recognition with particular emphasis on hand gestures and facial expressions, which outlines various models or tools often used for gesture recognition such as hidden Markov models, particle filtering and condensation, finite-state machines, optical flow, skin color and connectionist models.

As one of non-contact human-machine interaction techniques, head pose recognition has very wide-ranging applications, such as auxiliary driving systems

* This work is partly supported by the national Natural Science Foundation of China (60772063).

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 694–701, 2011.

[4][5], e-learning[6], meeting[7], visual user interface [8], virtual mouse driving [9], intelligent environments [10][11], Intelligent wheelchair systems [12] and robot systems [13][14].

Different from the above systems, this paper provides a head pose identification method using nose's relative position information in a face region and develops a head pose based game system, which can offer users an interesting and beneficial relaxation manner to prevent them from neck vertebral ankylosis, shoulder joint arthritis, xeroma and so on. The identified head poses are used to control a virtual robot walking in a virtual maze built with Microsoft Robotics Developer Studio.

2 Head Pose Recognition

Head poses are usually defined according to factual application's requirements. For example, in yes/no cases, only two head poses are needed. In our system, the virtual robot will be controlled to move in a virtual maze scene according to different head poses. The robot should be able to move forward/backward, turn left/right and stop. We define seven head poses for this application: UP, DOWN, LEFT, RIGHT, H-LEFT, H-RIGHT, and FRONT. UP means a raising head pose, DOWN is a lowering head pose, LEFT and RIGHT is a turning left and a turning right head pose of the real user respectively, H-LEFT and H-RIGHT is a slightly turning left and a slightly turning right head pose respectively, and FRONT is a looking forward pose. These head poses respectively correspond to the commands that let the robot go forward and backward, turn left, right, half-left and half-right, and stop.

Three Adaboost cascade classifiers [15] are prepared to detect a face in an image and the nose in the face region. The FRONT0 classifier is used to detect a non-profile face in the image and identify a head pose FRONT0. FRONT0 is an intermediate result which may be one of H-LEFT, H-RIGHT, FRONT, UP and DOWN. The LEFT classifier is used to detect a left profile face and identify the head pose LEFT. A nose classifier is used to detected the nose's position and size in an detected face image. The nose's position in the face region is used to decide the head pose FRONT0 is H-LEFT, H-RIGHT, FRONT, UP or DOWN.

2.1 Face Region Detection and Primary Identification

For this game application in most cases there is only one face in each image. The head region of the user in the image can be represented by its face region. The face detection method used in this paper is the same as in [15]. An integral image representation is also used to compute Haar-like features, and cascade classifiers are built using the AdaBoost learning algorithm. In this way we establish three classifiers to detect FRONT0 (non-profile face), LEFT (left profile face) and nose in a detected face region.

Haar features used mainly include three categories as in Fig. 1: edge features, linear features, central and diagonal features. These features can be defined as feature templates where there are two kinds of rectangles with white and black color respectively. The feature value of these templates is defined as the difference between the sum of the white pixels and the sum of the black pixels in the feature template.

Feature templates can be zoomed and arbitrarily placed in any sub-window of an image. Any zoomed form of a template is a feature. Finding these features and calculating their values are the foundation of training weak classifiers.

The integral image is an array where preserves the sum of pixels of every rectangular area bounded from the starting point to every image points. When we need the sum of pixels in a certain area, we can directly index the element of the array, no need to recalculate the sum again. Using such an integral Image, the computational time is the same for different features with various scales, and thus the detection speed can be greatly improved.

The features of profile faces are very different obviously from those of non-profile faces (including H-LEFT, H-RIGHT, FRONT, UP and DOWN), so we can firstly make a rough classification. The Adaboost algorithm is used to train the FRONT0 and LEFT classifiers, similar to the face detection method used in [15]. Training is such a process to select those with strong classifying ability from all features, form weak classifiers with those selected features, and build strong classifiers by using Adaboost cascade. The face and un-face sample images are firstly regulated to gray images with the size 20×20. For each sample $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $\boldsymbol{x}_i$ is a vector with 20 x 20 elements keeping all pixel values of one image, and $\boldsymbol{y}_i \in \{+1,-1\}$ denotes the class of the sample image. Therefore the input of a classifier is a 20 x 20 vector, and the output is a positive or negative value.

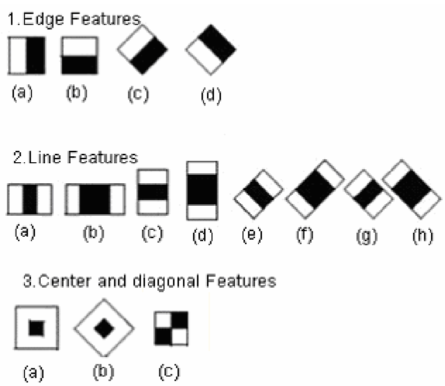


Fig. 1. Haar features including three categories: edge, linear, central and diagonal

We use the trained frontal and profile face classifiers in OpenCV to detect the face and roughly identify the head pose. The flow chart is shown in Fig. 2. The input of the FRONT0 classifier and the LEFT classifier is the whole image, and the output includes the position and the size of a face region and the identified result. If an image is neither a FRONT0 nor a LEFT through the two classifiers, then the image is flipped from left to right and the LEFT classifier is used again to identify if it is a RIGHT pose or no face in the image.

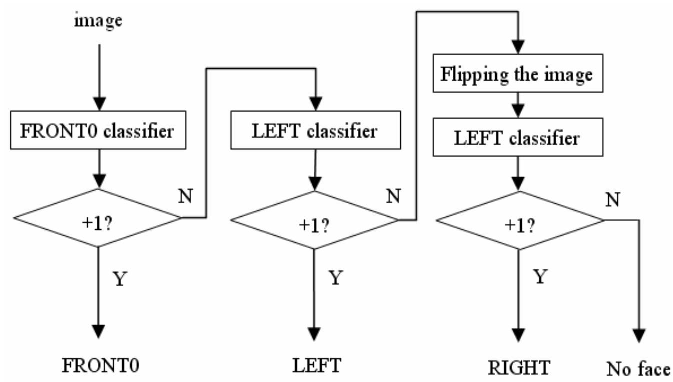


Fig. 2. The flow chart to identify FRONT0, LEFT, RIGHT, or no face in an image

2.2 Further Identification Using Nose's Relative Position

If the above result is FRONT0, then the head pose may be one of H-LEFT, H-RIGHT, FRONT, UP and DOWN. The nose's horizontal and vertical relative positions in the detected face region are used to decide the specific head pose from FRONT0. As we know, when people change their head poses, the facial feature in the image will change correspondingly, especially the geometry of the nose. For example, when the head turns from left to right, the nose's horizontal position in the face region will obviously moves towards the left (the right of the real user) in the images. When the head is raised down to up, the nose's vertical position will move up. According to this common sense, we use Adaboost algorithm again to train a nose classifier to detect the nose position in a detected face region.

In this application, a FRONT0 image is input to the nose classifier, and the output of the nose classifier includes a nose area's position and size information (shown in Fig. 3), including the left-up point coordinates (px, py), the width $lenx$, and the height $leny$. Then the center coordinates ($c1$, $c2$) of the nose area can be obtained by $c1=px+lenx/2$ and $c2=py+leny/2$. Suppose the face region's width and height is Fx and Fy respectively, then the nose's relative positions in the face region can be given by:

$$POSX=(px+lenx/2)/Fx$$

$$POSY=(py+leny/2)/Fy$$

$POSX$ and $POSY$ are the nose's horizontal and vertical relative positions in the detected face region.

When using the nose classifier to detect the nose area, more than one nose could be found in a face region because of noise. Therefore we propose a method to select the best one according to the features of the human's face. By analyzing the features of multi-face images, we conclude that the nose's vertical relative position $POSY$ is about 0.6, and the nose's relative width *LEN*=*lenx*/*Fx* changes in [1/6, 1/3]. Hence we propose a scoring scheme to select the best nose area with the highest score depending on the values of *POSY* and *LEN*. The selection steps are the following.

For all detected nose areas:

- Let *score* =100 ;
- Compute every vertical relative position and relative width:
  $POSY=(py+leny/2)/Fy$
  $LEN=lenx/Fx$;
- Modify the score according to the value of *POSY*:
  $score1 = score * (1-\|0.6-POSY\|/0.6)$
- Modify the score again according to the value of *LEN*:
  If $LEN<1/6$, $score2 = score1 * [1-5*(1/6-LEN)]$;
  If $LEN \in [1/6, 1/3]$, $score2 = score1$;
  If $LEN>1/3$, $score2 = score1 * [1-1.5* (LEN -1/3)]$

End for. Select the nose area with the biggest *score*2 as the best one.

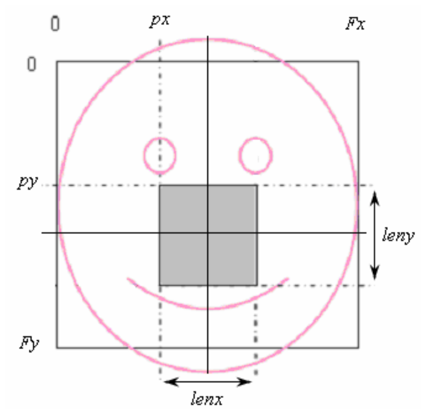


Fig. 3. Nose area's position and size the face region

The models for computing *score*1 and *score*2 are shown in Fig. 4.

From the best nose area, we can directly identify a FRONT0 head pose as H-LEFT, H-RIGHT, FRONT, UP, or DOWN, according to its center and the values of *POSX*和 *POSY*. The rules are described as follows:

1) If $c1=Fx/2$, the head pose FRONT, UP, or DOWN can be decided from *POSY*. If $POSY<0.55$, then the head pose is UP; If $POSY>0.65$, then the head pose is DOWN; If $0.55\leq POSY\leq 0.65$, then the head pose is FRONT.
2) If $c1\neq Fx/2$, the head pose H-LEFT or H-RIGHT can be decided from *POSX*. If $POSX<0.5$, then the real head pose is H-RIGHT; else it is H-LEFT.

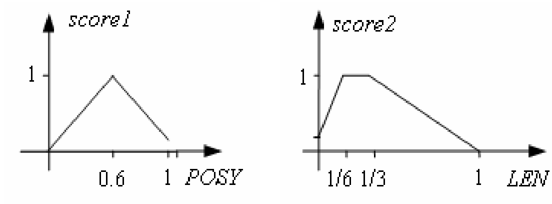


Fig. 4. The models for computing *score*1 and *score*2

3 Virtual Environment and Experiment Result

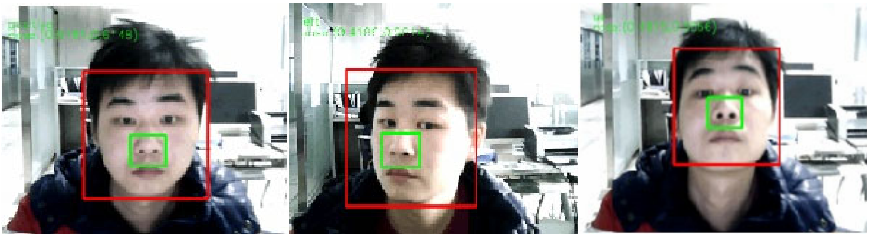
We construct a virtual robot and maze environment with Microsoft Robotics Developer Studio. The virtual robot based on the P3DX Pioneer robot model is controlled to move in the virtual maze environment by the real head poses of a user, with an anti-collision sensor in the front and back of the robot respectively. A virtual camera is also built to observe the continually variational maze environment, avoiding the user using a mouse to adjust the view points.

The virtual robot has two active front wheels and one passive back wheel, which can move by giving speeds on the two front wheels. Head poses and corresponding control commands are shown in Table 1. In the table, the function of SetDriveSpeed() is to set the speeds of the two front wheels, where the first argument is the speed of the left-front wheel, and the second argument is the speed of the right-front wheel. If the given speeds of the two front wheels are the same, the robot will go forward or backword; if the two speeds are different then the robot will turn around. For eaxmple, if the left speed is 0 and the right speed is 0.15 then the robot will turn left. When the robot executing the turning commands, the turning angle of LEFT is bigger than that of H-LEFT, and similarly the turning angle of RIGHT is bigger than that of H-RIGHT. If the detected head pose is FRONT or no face in the image then the robot will keep stopping.

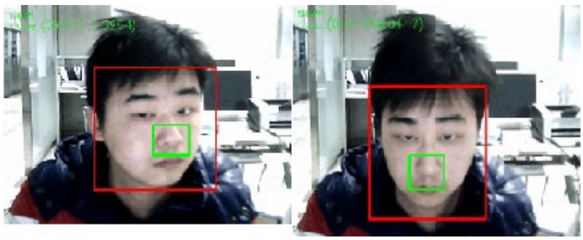
Table 1. Head poses and corresponding control commands

Head poses	Control commands
FRONT	SetDriveSpeed(0.0f, 0.0f)
UP	SetDriveSpeed(0.2f, 0.2f)
DOWN	SetDriveSpeed(0.1f, 0.1f)
H-LEFT	SetDriveSpeed(0.0f, 0.1f)
H-RIGHT	SetDriveSpeed(0.1f, 0.0f)
LEFT	SetDriveSpeed(0.0f, 0.15f)
RIGHT	SetDriveSpeed(0.15f, 0.0f)

In the experiments, a user's face image is input to the face classifiers to recognize a head pose FRONT0, LEFT or RIGHT firstly, and get the face region. If it is FRONT0, then the detected face region is further input to the nose classifier, and the nose's relative position and size can be acquired, from which the head pose can be determined as H-LEFT, H-RIGHT, FRONT, UP or DOWN. Fig. 5 shows several detected head poses, and Fig. 6 is an image sequence of robot moving.



(a) FRONT (d) H-LEFT (b) UP



(e) H-RIGHT (c) DOWN

Fig. 5. Five detected head poses. Here H-LEFT and H-RIGHT denote the real head poses of the user

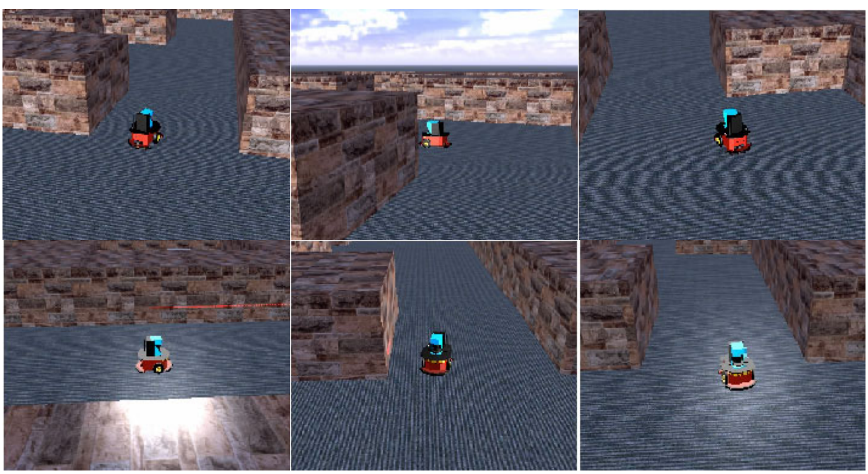


Fig. 6. A sequence of robot moving

4 Conclusion

This paper presents a game system based on the head pose identification techniques. The different head poses are used to control a virtual robot walking in a virtual maze which is built with Microsoft Robotics Developer Studio. The system is simple to operate not needing any touched input devices such as a keyboards, mouse or game manipulating bar except a cheap camera. This game can help people exercise their necks, relax their bodies and amuse themselves.

The system firstly uses Adaboost classifiers to detect a human face in the image and roughly identify the head pose, and then the nose's relative position information in the face region is used to further determine the head pose. Seven head poses correspond to seven different control commands to control the virtual robot moving.

In the future we will try to improve algorithms to improve the system's response speed and use head pose identification techniques to control a real robot.

References

1. Murphy-Chutorian, E., Trivedi, M.: Head Pose Estimation in Computer Vision: A Survey. IEEE Trans. on Pattern Analysis and Machine Intelligence 31(4), 607–626 (2009)
2. Ghorbel, M., Baklouti, M., Couvet, S.: 3D Head Pose Estimation and Tracking Using Particle Filtering and ICP Algorithm. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2010. LNCS, vol. 6169, pp. 224–237. Springer, Heidelberg (2010)
3. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. IEEE Trans. on Systems, Man, and Cybernetics–Part C: Applications and Reviews 37(3), 311–324 (2007)
4. Tran, C., Trivedi, M.: Driver Assistance for Keeping Hands on the Wheel and Eyes on the Road. In: Proc. of The Int. Conf. on Vehicular Electronics and Safety, pp. 97–101 (2009)
5. Murphy-Chutorian, E., Trivedi, M.: Head Pose Estimation and Augmented Reality Tracking: An Integrated System and Evaluation for Monitoring Driver Awareness. IEEE Trans. in Intelligent Transportation Systems 11(2), 300–311 (2010)
6. Asteriadis, S., Tzouveli, P., Karpouzis, K., Kollias, S.: Estimation of Behavioral User State Based on Eye Gaze and Head Pose—Application in an e-learning Environment. Multimedia Tools Appl. 41, 469–493 (2009)
7. Ba, S.O., Odobez, J.M.: Recognizing Visual Focus of Attention from Head Pose in Natural Meetings. IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics 39(1), 16–33 (2009)
8. Morency, L.P., Sidner, C., Lee, C., Darrell, T.: Head Gestures for Perceptual Interfaces: The Role of Context in Improving Recognition. Artificial Intelligence 171(8-9), 568–585 (2007)
9. Fu, Y., Huang, T.S.: hMouse: Head Tracking Driven Virtual Computer Mouse. In: Proc. IEEE Workshop Applications of Computer Vision, pp. 30–35 (2007)
10. Pentland, A., Choudhury, T.: Face Recognition for Smart Environments. Computer 33(2), 50–55 (2000)
11. Trivedi, M.: Human Movement Capture and Analysis in Intelligent Environments. Machine Vision and Applications 14(4), 215–217 (2003)
12. Jia, P., Hu, H., Yuan, K.: Head Gesture based Control of an Intelligent Wheelchair. Industrial Robot. 34(1), 60–78 (2007)
13. Matsumoto, Y., Sasao, N., Suenaga, T.: 3D Model-based 6-DOF Head Tracking by a Single Camera for Human-Robot Interaction. In: 2009 IEEE Int. Conf. on Robotics and Automation, Japan, pp. 3194–3199 (2009)
14. Wallhoffa, F., Rehrla, T., Mayerb, C.: Real Time Face and Gesture Analysis for Human Robot Interaction. In: Proc. of SPIE Real-Time Image and Video Processing, vol. 7724, pp. 77240A-1–77240A-10 (2010)
15. Viola, P., Jones, M.J.: Robust Real-time Detection. Int. J. of Computer Vision 57(2), 137–154 (2004)

Author Index