Aaron Marcus (Ed.)

LNCS 6770

Design, User Experience, and Usability

Theory, Methods, Tools and Practice

First International Conference, DUXU 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 2011, Proceedings, Part II

2 Part II

HCI2011
INTERNATIONAL

Springer

Lecture Notes in Computer Science 6770

Aaron Marcus (Ed.)

Design, User Experience and Usability

Theory, Methods, Tools and Practice

First International Conference, DUXU 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 9-14, 2011
Proceedings, Part II

Springer

Volume Editor

Aaron Marcus
Aaron Marcus and Associates, Inc.
1196 Euclid Avenue, Suite 1F
Berkeley, CA, 94708-1640, USA
E-mail: Aaron.Marcus@AMandA.com

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21707-4 e-ISBN 978-3-642-21708-1
DOI 10.1007/978-3-642-21708-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929078

CR Subject Classification (1998): H.5, H.3-4, K.6, D.2, C.2, K.4

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 Aaron Marcus, contains papers in the thematic area of design, user experience, and usability (DUXU), addressing the following major topics:

- DUXU in Web environment
- DUXU and ubiquitous interaction/appearance
- DUXU in the development and usage lifecycle
- DUXU evaluation
- DUXU beyond usability: culture, branding, and emotions

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 4, LNCS 6764, Human–Computer Interaction—Users and Applications (Part IV), edited by Julie A. Jacko
- Volume 5, LNCS 6765, Universal Access in Human–Computer Interaction—Design for All and eInclusion (Part I), edited by Constantine Stephanidis

- Volume 6, LNCS 6766, Universal Access in Human–Computer Interaction—Users Diversity (Part II), edited by Constantine Stephanidis
- Volume 7, LNCS 6767, Universal Access in Human–Computer Interaction—Context Diversity (Part III), edited by Constantine Stephanidis
- Volume 8, LNCS 6768, Universal Access in Human–Computer Interaction—Applications and Services (Part IV), edited by Constantine Stephanidis
- Volume 9, LNCS 6769, Design, User Experience, and Usability—Theory, Methods, Tools and Practice (Part I), edited by Aaron Marcus
- Volume 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 II

Part I: DUXU in Web Environment

Part II: DUXU and Ubiquitous Interaction / Appearance

Part III: DUXU in the Development and Usage Lifecycle

Part IV: DUXU Evaluation

Part V: DUXU beyond Usability: Culture, Branding, and Emotions

Table of Contents – Part I

Part I: DUXU Theory, Methods and Tools

Part II: DUXU Guidelines and Standards

Part III: Novel DUXU: Devices and Their User Interfaces

Part IV: DUXU in Industry

Part V: DUXU in the Mobile and Vehicle Context

Challenges and Opportunities of Hotel Online Booking in China

Wei Ding

Mariott International, Inc.,
10400 Fernwood Road, Bethesda, MD 20854, USA
Wei.ding@marriott.com

Abstract. This paper provides insights into Chinese consumers' behavior, attitude, and preference for travel planning and research, hotel selection and on-property spending patterns. Challenges and opportunities are discussed and practical recommendations are made for global hotel companies to create and execute their multi-channel eCommerce strategy by focusing on culturally savvy website localization and wise online marketing.

Keywords: eCommerce in China, website localization, online travel booking, consumer behavior, hospitality.

1 Introduction

China, as the 2nd largest economy with 457 million Internet and 309 million mobile Internet population, owns 3000+ travel websites. While the penetration rate of online travel booking is still low, an explosive growth in this sector is expected to take place in the next few years [1].

With almost all the world's major hotel companies entering the China market, it has become an urgent topic to understand the China's social and cultural context, the market and the consumer in the online hotel booking landscape. While there are commonalities between the western and China markets, the focus of the discussion here is on the uniqueness of the culture and the corresponding consumer behavior and expectation. It aims to provide practical guidance for global hoteliers to best align their products and services with the China market and therefore to achieve win-win outcomes. Note this paper is more focused on hotel online research and booking by transient customers instead of group intermediaries for meetings and events.

2 Social and Cultural Context and Online User Behavior

2.1 Cultural Characteristics

Relevant to our discussion here, China's culture can be characterized by two of the dimensions in Hofstede's model: collectivism and high power distance[5]. Based on Hall's culture style theory, China's culture is considered as "high context [2].

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 3–12, 2011.

China is a typical *collectivism* culture, where people are integrated into strong cohesive in-groups. They tend to conform to group norms. Under many circumstances, the opinion of the collective will influence and even change the individual's own view. Therefore, "word of mouth" is an important influencer of consumer behavior. People favor "best sellers" and tend to follow the crowd. Thus, in general, market leaders have the best opportunity to become even stronger.

Because of the nature of collectivism, relationships are extremely important. Social and personal networks drive achievement. This is applicable to both individuals and businesses. Companies allied with other highly performing companies are generally perceived as well established and trustworthy. Friendship links to other companies and businesses are commonly seen on many highly respected brand sites in China. This also implies that global companies should consider establishing good business partnership with strong local brands.

Power distance is the extent to which individuals accept and expect the unequal distribution of power. China had a higher score than the US in power distance according to Hostede's studies. People respect authority, seniority and celebrities, tend to focus on prestige, credentials, certifications and awards and use them as indicators of quality assurance. For global companies whose brands and identities are unfamiliar to the Chinese customers, leveraging the abovementioned information for marketing and promotions can be quite effective.

In *high context* (HC) cultures, the communication is identified as indirect, ambiguous, maintaining of harmony, reserved and understated. Studies show that websites in HC culture tend to use more imagery and animated effects than those in low context culture; Websites are expected to be diverse and sophisticated; Navigation paths are more subtle and less linear.

2.2 Cultural Preferences

China has a very long history and profound culture with many unique aspects Relevant to the topic of our discussion, here are the highlights:

Every name has a meaning

In the Chinese culture, almost all names have some meaning, especially when it comes to business names. International companies entering the China market need to pay special attention to the way their brand is translated. A good translation often needs to meet the following criteria:

- phonetically sounds similar to its original pronunciation,
- sounds pretty and carries elegance,
- the meaning is relevant to the customer and matches the company's business value,
- easy to remember, and
- there should not be other (negative) interpretations of the name, even based on homophones.

Coke Cola's name translation changed from sounding like "biting the wax tadpole" to "putting the mouth to rejoice", made a huge difference in impacting customer's perception of the brand although both phonetically sound similar. IKEA's Chinese translation "perfect for the home" is also highly regarded. On the other hand, Best

Buy recently shut down all their stores in China. Its name translation " , could be interpreted as "always thinking of buying (from the store)", or "thinking a hundred times before buying". The latter could sound negative to the business.

Symbolic meanings are important

As mentioned earlier, in the high context culture, people tend to use more images and icons to communicate to avoid over-straightforwardness. In the Chinese culture, certain colors, numbers and animals are associated with strong symbolic meanings as shown in Table 1. Some of the associations are based on ancient beliefs, and many are simply derived from the homophones. For example, "four" is a homophone of "death". It is very important to use the appropriate images in the right context.

Table 1. Symbolic Meanings of Colors, Numbers and Images of Animals

Category		**Meaning**
Colors	Red	happiness and good luck; or financial gains
	Green	Financial losses;
	Gold	Royal and power
	White	Death
Numbers	Eight	Homophone of "Prosperous"
	Three,six, nine	Lucky
	Four	Homophone of "death", bad luck
Animals	Dragon	Noble & powerful
	Tiger	Powerful
	Dog	Loyal
	Cat	Clever but disloyal
	Donkey	Stupid

The Chinese language is complex and dynamic

Being intricate and complex, the Chinese language has its own analogies and expressions. When it comes to translation, special attention needs to be paid to the language nuances. For example, if the hotel room promotion is 20% off the regular price, it should be stated as 80% offers in Chinese. Also the address in China goes from the broad to the specific, which is opposite to the western format.

In addition, the separation between Mainland China and Taiwan due to historical reasons makes the language even more complex: traditional Chinese characters are used in Taiwan while simplified in China; Many terms and concepts originated from the West also have different translations for China and Taiwan.

Recently, there have been a number of "Internet terms and phrases" added to Chinese people's vocabulary. For example, microblog is also called "scarf" because of the homophonic relationship between the terms. "Fans (of celebrities)" is now translated into "thin vermicelli". To make connections with the target customer, global companies need to beware of these expressions.

2.3 Hotel Information Searching Behavior

Overall, the Chinese Internet population is younger than those in other countries. 81% of them are younger than 40 years old, and 53% are aged between 20-40.

According to iResearch's data[3], an increasing number of Chinese users use the Internet to research on hotel information year over year. In 2010, the top three popular channels are: general search engines, OTAs and portal sites. For travel information research, such as travel ideas and price comparison, direct hotel sites are used less than search engines and OTAs, largely because during the research phase, 70% of the users had no idea about which hotel to stay [7].

For search engine activities, after the destination is determined, frequent keywords used fall into two categories: brand names and well-known landmarks. iResearch's findings [4] show that high-end international brands are less searched than economy hotel brands. There are two major reasons behind this: first, the economy hotel market is much larger than the high-end hotel market, and the target audience of economy hotels is much more price sensitive than the high end hotel audience. Secondly, the price sensitive users are much more likely to conduct online research than the less price sensitive users.

The landmark keywords often include attractions, shopping areas, traffic hubs, and entertainment centers. When selecting keywords for SEO purposes, hotel websites should make good use of the relevant landmark names.

2.4 Trip Planning Timeline and Hotel Reservation Behavior

According to the iResearch data [3], leisure travelers do not take a long time for planning compared to their western counterparts: 80% of the leisure trips were planned within half a month before departure and 32% of the reservations were made after arriving at the destination. This suggests that last minute deals via email or IM may be attractive to the right customers and mobile can play a big role in same day hotel booking. Location based marketing message may work well once the related technology gets more robust in China.

When it comes to the actual hotel booking[3], although OTA sites still lead the way, which account for 32% of the total bookings, direct hotel channels (brand site plus phone and mobile reservations) are now #2, claiming 26% of the market share, followed by online shopping platforms (such as Taobao) (11%). On the other hand, there are still a large number of internet users who end up booking offline (by calling to or walking in to the hotel property).

2.5 Key Hotel Selection Drivers

Similar to the leisure travelers in the US, the #1 factor driving the Chinese travelers' hotel selection decision is price. Convenience of the hotel location is the second most popular reason. The subsequent influencers are amenities, hotel reviews and the star-ranking of the hotel [3]. For travelers especially those who have no idea or are unsure about which hotel to stay, hotel reviews play an important role in their decision making. Recent studies [8] show that 81-90% of online shoppers or online travel bookers read product/hotel reviews before purchases.

Table 2. What hotel customers care most?

Factors impacting hotel reviews	Leisure Travelers	Business Travelers
#1	Price	Amenities
#2	Location	Transportation
#3	Transportation	Service quality
#4	Amenities	Location
#5	Service quality	Food & beverage
#6	Food & beverage	Price

Based on a comprehensive analysis of customer-contributed hotel reviews, Qunar, one of the top Chinese vertical travel search engines, reported [8] that the most important factors the hotel guest cares about differ between leisure travel and business travel (as shown in Table 2). Their results show that business travelers' top priority is hotel amenities while leisure travelers are much more price sensitive and look for the best performance/cost ratio.

Economy hotel customers more focus on whether the property is brand new or has been renovated followed by high speed internet and breakfast while customers of 3-star or lower hotels pay more attention to breakfast and business center.

2.6 On-Property Consumer Behavior Patterns

According to Goldman Sachs' research, non-room revenue (e.g., food and beverage) typically makes up 40% of the hotels revenue and in some cases over 50% [6]. The (high-end) hotel is generally perceived as the location for frequent business travelers to host clients. Luxury is seen as crucial for self-marketing & as a "business investment". Public space is perceived more important than the guest room.

In terms of the luxury tier loyalty program, some studies show that people are more interested in earning points for status add-ons than free nights and other studies [6].

3 Characteristics of China's Hotel Online Market

China's hotel online market has its own unique characteristics. First, as discussed earlier, the penetration rate of online hotel booking is still low. Most travelers use online channels for research and the reservation takes place offline. Secondly, global hotel companies face tough competition from local hotel chains, OTAs and other third party online travel platforms. Third, global hotel companies need to quickly get adapted to the local culture in order to replicate their eCommerce successes with other markets. Detailed discussions are as follows:

3.1 Tough Competition in the Hotel Online Research and Booking Space

Just like in the western markets, hotel companies face tough competition with many other players, such as online travel agencies (OTAs), online platform providers (e.g., Taobao), travel and general search engines (e.g., Qunar and Baidu), and travel community sites (e.g., Daodao). Next we are going to look at them more closely.

Dominance of OTAs

In general, there is a hate and love relationship between the hotel companies and the OTAs. According to many case studies, the OTA channel could be 10 times more expensive than the hotel's direct online channel [12]. However, when travel supply outweighs demand, travelers are shopping around and hoteliers are more susceptible to discounting and working with the OTAs. In the past two years during the economic downturn, that happened to the top 30 western hotel companies [12]. Nevertheless, the direct channel revenue for these companies is still twice as that from the OTAs.

In China, the situation is not quite the same yet. While chained hotels with good brand awareness are more capable of expanding a membership network and build up a central reservation system (CRS), independent hotels have to rely on the OTAs to sell their rooms. With about 90% of the hotels being independent [11], the OTAs become a natural choice for customers to research and book hotels.

As of early 2010, Ctrip had 55% of the OTA market share, followed by eLong (22%), one of Expedia's China components. Ctrip has well integrated its online resources and traditional travel agency functions via strategic acquisitions. It has the largest call center in the world. 40% of its revenue comes from online bookings while 60% from phones. Its business model seems well suited to the China market: most users tend to use online channels for research and planning while making reservations offline. 80% of Ctrip's hotel revenue comes from 4-star and above hotels.

eLong is currently more focused on online hotel booking, especially for hotels of mid-tier or below.

Other 3rd party channels

The vertical travel search engines and travel community sites (e.g., Daodao) also attract a large amount of hotel research traffic. Qunar and Kuxun are the leading vertical travel search engines. By traffic, Qunar is the 2nd largest travel site in China. Daodao is now the China branch of Tripadvisor and hosts a large volume of user generated reviews.

Taobao: the rising online platform

More recently, China's largest eCommerce player Taobao also entered the travel market. With its huge user base of 370 mln registered accounts, it has a great potential to also play a large role in online hotel booking.

With these various channels available to leverage and yet compete with, hotel companies need to have a clear understanding about who and where their customers are, how they can be best reached, and what channels are most suitable for the business.

Online marketing spending pattern

The most recent data show that hotel companies' online marketing spending is primarily on portal websites, accounts for 73% of the total spending [3], followed in turn by lifestyle (including travel) sites, finance sites, video sites and news sites. Among the portal sites, the MSN portal claims the largest share of 57% of hotel advertisement costs.

3.2 Special Challenges for Global Hotel Companies

While global hotel companies bring in cutting edge management expertise and admirable services to the market, they are also facing additional challenges when executing their ecommerce strategies in China:

Different SEO practices
In China, search engine penetration is 95% and the dominant search engine is Baidu , which is also the number one source for hotel information research. More than 20% of the ctrip traffic comes from Baidu. An eyetracking research [14] has shown that when examining search results, Baidu users don't follow the same scan patterns as the Google users (focusing on the upper left "golden triangle" area), and the sponsor links and organic results are also treated differently. Therefore, SEO strategies and rationale tied to Google may not completely work with Baidu.

Brand differentiation
Although the general attitude of Chinese customers towards the high-end western hotel brands is positive, people tend to treat them as in the same group because of the unfamiliarity of the brands. Certain companies enjoyed the first movers' advantages, but since the competition is getting so intense, all companies need to figure out the most effective way to foster customer loyalty and increase brand recognition in order to stand out from the pack.

Website localization and cultural differences
Using direct online channels to drive revenue and build customer loyalty is the ultimate goal for large hotel companies, and website localization is a big topic for those that are expanding their business in China. A popular approach has been to design one website and make it adapted to various languages and regions without engineering changes. This helps the company maintain centralized control over the content dissemination and keeps the development costs reasonably low. However, this does not warrant the same level of customer satisfaction, effectiveness and profitability for a different market. In order to make the website culturally relevant, an array of factors need to be carefully thought through, including user expectations, cultural preferences, business practices, language nuances, related government policy and many more. Great lessons can be learned from some world renowned companies' successful and unsuccessful stories in China, such as eBay [9]. Also, regional offices should be enabled to make necessary dynamic changes to the site in a timely manner.

Target customers
A large portion of the target customers for these hotels are business executives, who are less price sensitive and less likely to do hotel research or book hotels themselves. Instead, dedicated travel staff or assistants take care of their travel. Understanding who, how and when are involved in the different stages of the travel journey (from research & planning, hotel booking, pre-arrival, staying on property to post-trip) for this type of customers will help hotels better strategize personalized marketing, online product development, on-property services and loyalty program operations.

4 Opportunities and Recommendations for Global Hotel Companies

Compared to the Western hotel markets, the China has the following differences:

- The percentage of chained hotels is much smaller (7%);
- Hotel online booking penetration is still very low (8%);
- Most global hotel companies are competing for the upper mid-tier and upper tier market;
- OTAs are dominating online hotel booking;
- The leading search engine Baidu works differently from Google;
- The Taobao B2C platform is getting an increasingly higher participation rate from all kinds of high-end consumer brands.

These differences bring challenges and opportunities for those global hotel companies. Specific recommendations are as follows:

4.1 Maximizing the Direct Online Channel

It is evident that the direct online channel has the most advantages. It allows the brand to interact with and engage its customer directly throughout different phases of the travel journey, grow brand loyalty and build long term competitive advantages.

In recent years, branded economy hotels have grown their direct online channels effectively [11]. Seven Days Inn for example has cut out its indirect online channels. With 80% of its customers as returning customers, 7 Days Inn has 70% of the reservations coming from direct Internet channels [4]. Other economy hotel chains are also able to keep their OTA channels under 8% [13]. The increased visibility of the direct online channels gives the customer more options and help them better understand the advantages of these channels. It also creates the momentum for hotels at all levels to invest more in their own direct channels, reduce costs, and minimize dependencies on expensive channels.

Most of the global companies, managing 4 or 5-star hotels in China, have robust online distribution networks and have been successful in their niche markets. Their brands are perceived prestigious and desirable in China especially thanks to the quality of the services, exclusive locations, grand facilities and amenities. In order for their direct channels to be better recognized by the Chinese customers, they need to do the following:

- Create a culturally savvy online marketing strategy to promote their brands and attract traffic.

As mentioned earlier, in a collectivism culture with relatively high power distance score, "word of mouth", thought leaders, professional experts and celebrities have large influence on people's behavior, and also people pay attention to credentials and certificates. Global companies should make good use of related information, such as celebrity usage, company awards, accolades, rankings and prestige as well as "digital word of mouth" information (e.g., hotel reviews), and integrate them into their SEO, SEM and online advertising practices to increase brand awareness and attract traffic to the website.

In addition, hotel companies need to focus on the advantages and exclusive benefits of the direct online channel, such as the promise of the brand, availability of room upgrades, hotel packages, and flexible cancellation policy.

- Optimize the localization of the official Chinese website to increase website stickiness and traffic conversion rate.

Many global hotel companies have launched their website in Chinese, and most of them use pictures of Chinese people and showcase local architectural achievements to display local promotions and make connections with the customer. This makes the site feel culturally relevant and is well aligned with the values of China's high context and collectivism culture, which is a great starting point.

To bring these websites to the next level, the companies need to strengthen their Chinese language processing capabilities so that the user is not forced to enter Pinyin for their name and address. Also they need to deepen the level of localization, which requires a lot of detail-oriented work. Table 3 below shows some subtle differences between the western websites and Chinese websites.

Table 3. Differences between Typical Western and Chinese Hotel Websites

Subject	**Western Site**	**Chinese Site**
Guest information required	Email & address	Email & cell phone number
Room guarantee	Credit card	Optional
Online payment method	Credit card	3rd party payment, bank card and bank transfer

Given that Chinese users have a slightly different perception of site usability: they tend to place more emphasis on the visual appearance and prefer pictures and photos over text, hotels should consider displaying more large-sized, emotionally engaging photos to show the grander of the hotel's architecture, lobby and public areas within the building. In addition, the website of high-end hotels needs to generate the right impression for the customer that matches the perception of the brand.

Global companies should also get strategic in introducing customers to their large distribution network around the world. When China's outbound travel to the West gets more common, the customer's familiarity and loyalty to the western brands will influence their hotel and travel decision in a positive way.

4.2 Craft Out a Multiple Online Channel Strategy to Sustain the Growth

As the Internet evolves to be more and more open, social, mobile and intertwined, the available online channels become more sophisticated and diversified. In addition to the direct online channel, strategies for SEO, OTAs, virtual malls, and mobile and social media marketing need to be carefully thought out as well. Having multiple channels complement each other will make the company's growth more sustainable and less immune to the economy situation, and thus out-smart the competition.

Because of the collectivism culture in China, typically existing market leaders tend to be more likely to get stronger. Identifying the market leaders and partnering up with the right players is critical to any global hoteliers. Since the China market is so dynamic and still fast developing, new players and new business models may emerge any time. Therefore, it is also equally important to keep monitoring the market and making strategy adjustments accordingly.

Overall, the companies that are culturally savvy and capable of aligning their strengths with the customers' needs, will stand out from the "western group", earn Chinese customers' loyalty, and eventually become successful in the China market.

References

1. CNNIC (China Internet Network Information Center). China Internet Status Report (2011), http://www.cnnic.cn/
2. Hall, E.T.: Context and meaning. In: Samovar, L.A., Porter, R.E. (eds.) Intercultural Communication: A Reader, 9th edn., pp. 34–43. Wadsworth Publishing Co, Belmont (2000)
3. iResearch. China Airline & Hotel Internet sales channels Research Report (2010)
4. iResearch. China online travel research report (2008 -2009) via eNet, http://www.enet.com.cn/article/2009/0603/A20090603481738.shtml
5. Hofstede, G.: Culture's consequences: comparing values, behaviors, institutions and organizations across nations, 2nd edn. Sage Publications, Inc., Thousand Oaks (2001)
6. Sachs, G.: Global investment research report on hotel companies in China (June 2010), http://www.gs.com
7. Epsilon. Price, location and reviews drive hotel selection process (October 2010), http://www.hotelmarketing.com/index.php/content/article/price_location_and_reviews_drive_hotel_selection_process/
8. Qunar. Trends of hotel reviews in China (2010), http://www.chinaleisure.org/mediafile/201005/11/P201005111434267688794303.pdf
9. So, S. and Westland, J.C. How eBay lost the China market (2009), http://news.alibaba.com/article/detail/business-in-china/100151761-1-how-ebay-lost-china-market.html
10. ComScore. Top three leaders of online travel in China (2010), http://news.youth.cn/gd/201010/t20101021_1376720.htm
11. Morgan Stanley. China Leisure and Lodging (2010), http://www.morganstanley.com
12. Starkov, M. The Good and (Very) Bad News in the Online Distribution Channel (2010), http://www.hospitalityebusiness.com/blog/the-good-and-very-bad-news-in-the-online-distribution-channel/
13. Tnooz. Travel technology trends at Global Tourism & Travel Summit (2010), http://www.tnooz.com/tag/7-days-inn/
14. Hotchkiss, G. Chinese Eye Tracking Study: Baidu Vs Google (2007), http://searchengineland.com/chinese-eye-tracking-study-baidu-vs-google-11477

Analysis of Causal Relationships between Blog Design Criteria

Chun-Cheng Hsu

Assistant professor, Department of Communication and Technology,
National Chiao Tung University, 1001 Ta-Hsueh Rd., Hsinchu 300, Taiwan
chuncheng@mail.nctu.edu.tw

Abstract. There are numerous excellent studies devoted to blog design; however few of these can explain the interaction between evaluation criteria in a systematic way. The purpose of this study is to explore the causal relationships between the criteria for blog design. Since design is a multiple criteria decision-making problem, this study uses the Decision Making Trial and Evaluation Laboratory method (DEMATEL). The DEMATEL method is used to simplify and visualize the interrelationships between criteria in making a decision. This study adopted seven important criteria that influence blog design: Aesthetic Layout, Multiple Layout Style Choice, Ease of Management, Ease of Registration, Storage Capacity, System Stability, Friendliness to Beginners. According to the results of DEMATEL analysis, the impact-relations map was obtained. The author hopes that this study will make a useful contribution to better understanding blog design.

Keywords: Blog design, Multiple Criteria Decision-Making, DEMATEL method.

1 Introduction

This study aims to integrate design criteria into a decision-making cause and effect model. A thorough literature review is conducted, mainly focusing around two central themes. The first theme involves studies on webpage and blog design, while the second involves the research in the field of decision-making.

1.1 Web and Blog Design Evaluation

Numerous research efforts have explored this multifaceted discipline, proposing research methods that can be divided into several types. The simplest method is to conduct a quality evaluation survey such as a service satisfaction survey while a more complex research method would involve utilizing the proposed models to scrutinize the relationships between variables; such a method would probably apply the Technology Acceptance Model (TAM). Guo & Salvendy [1] studied Chinese online shopping websites, using a survey to pinpoint 15 key elements that determined users' degree of satisfaction. These elements included quality content, service content, and appearance

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 13–19, 2011.

description. Through identifying these key elements, they sought to assist web designers in preparing content for e-business websites that enhances the user's experience. Nielsen and Tahir [2] approached the problem from the usability perspective, and suggested 113 criteria for webpage design.

Furthermore, in considering the application of the Technology Acceptance Model (TAM) in explaining intranet usage, Horton et al. [3] found that factors such as usefulness, ease of use, and intention determined user acceptance. Lee and Kim [4] went further by including external factors in examining which elements influence users' acceptance of a website. They identified external factors such as technical support, web experience, task equivocality, and the perceived ease of use as influencing intranet use.

Numerous studies have paid much attention to webpage design, but in comparison, studies on blog design and its visual elements are lagging far behind. Existing researches such as Guan and Liu [5] have evaluated the usability of blog designs and advised the urgent need for improvement. Hsu and Lin [6] affirmed the importance of technology and design in blogs, adding that the degree to which bloggers interact is positively associated with the enjoyment they derive from using appropriately designed interfaces. Such intrinsic motivation is an important consideration.

In order to narrow the scope of this research, it is necessary to restrict it solely to personal blogs to prevent ambiguity in its findings. It explores which key design factors bloggers believe to be important in blog design. On one hand it refers to evaluation criteria for web page design, while on the other it applies DEMATEL for the first time to analyze blog design.

1.2 Decision Making Trial and Evaluation Laboratory Method

In recent years, many scholars have favored the application of a multiple criteria decision-making (MCDM) approach to problem-solving in such disciplines as social science. This is to overcome inherent limitations that are inevitable in single method models. The multifaceted approach provided by MCDM can often help identify ingenious resolutions to complex problems and offer ample explanations and suggestions. For example, Tzeng, Chang, & Li [7] applied the hybrid MCDM model to analyze the effectiveness of e-learning programs. This hybrid MCDM model they used was based on factor analysis, DEMATEL, fuzzy theory, and the analytic hierarchy process (AHP) method. In concluding their study, the interrelationships between independent criteria and the weights of these criteria were identified, providing developers with a useful reference for the design of e-learning program websites. Chen & Chen [8] developed an innovation support system for evaluating the operative performance of Taiwan's tertiary education institutions. They used a methodology based on DEMATEL, the fuzzy analytical network process (FANP) and the technique for order preference by similarity to an ideal solution (TOPSIS). The results of the analysis served as benchmarks guiding these institutions towards maximizing innovation and creativity in structuring better education systems.

DEMATEL was developed in 1971 by the Science and Human Affairs Program at the Battelle Memorial Institute's Geneva research center [9]. It has been widely applied to complex problems in scientific disciplines such as technology, environmental science, and anthropology. Conventional quantitative analyses rely more on statistical analysis and applications of linear structure models. A good example of this is factor analysis, which extrapolates the degree of user satisfaction from surveys and interviews [7]. Many prior studies focusing on blogs did not employ methodologies that comprehensively examined the interrelationships between influencing factors, and ignored the interrelationships between criteria. DEMATEL, however, is able to compensate for these insufficiencies. It allows the relationships between evaluation criteria to be determined and a value structure established. DEMATEL is useful in applications in which appropriate decisions must be made.

Liou et al. [10] proposed a hybrid model that combined ANP with DEMATEL to extrapolate the weighted significance of organizational and managerial factors in the evaluation of airline safety management systems. This study found that although traditional airline safety management systems emphasize incident investigation and analysis, the key determinants that influence the management of airline safety are in fact strategic and policy factors. Similar research conducted by Wu [11] also proposed a hybrid model that could be used to analyze how a business could effectively utilize and transform knowledge into a competitive advantage, and also help it evaluate and choose better information management strategies.

In recent years, DEMATEL has been much praised for its invaluable application in disciplines such as social science and the realm of decision-making. In practice decision makers face both decisions that have independent criteria and those that have interacting criteria. Thus, isolating which criteria are key is beneficial in practice. However as DEMATEL has rarely been used as an analytical tool in the design field, this research seeks to reveal the potential applicability of DEMATEL to blog design and development by proposing a hybrid model that combines factor analysis with DEMATEL, in order to identify the key factors and their interrelationships of criteria within each factor.

2 Method

This research seeks to apply DEMATEL to gain insight into the causal relationships between the identified criteria and their degrees of influence. The three basic assumptions of DEMATEL are: (1) clarity in setting research questions: at the research planning stage, researchers must ensure that the research questions they set are clear; (2) clear association in the relatedness between criteria: the weighted association between criteria of the research question must be indicated by allocating them rankings in magnitude; (3) understanding of the characteristics of each factor arising out of the research question followed by supplementary conclusions after analysis [7] [11].

DEMATEL analysis was used to identify the interrelationships between the criteria. According to Hsu [12], there are seven important criteria that influence blog design: Aesthetic Layout(C1), Multiple Layout Style Choice(C2), Ease of Management(C3), Ease of Registration(C4), Storage Capacity(C5), System Stability(C6), Friendliness to Beginners(C7). This study adopted and incorporated these criteria into a questionnaire, which was then given to 30 experts. These experts included experienced blog users and website-related scholars. The experts ranked each criterion using the Likert 5 point scale.

3 Results

The purpose of this study is to find the relationships between these design criteria. Therefore, DEMATEL analysis was used to calculate the weighted significance of each criterion based on the interrelationships shown on the impact-relations map.

3.1 Total Relation Matrix and Causal Influence Table

Based on the responses of 30 experts, the total relation matrixes were calculated (Table 1).

Table 1. The total relation matrix

	C1	C2	C3	C4	C5	C6	C7
C1	0.366	0.557	0.507	0.461	0.422	0.46	0.6
C2	0.562	0.446	0.571	0.517	0.473	0.512	0.687
C3	0.504	0.548	0.441	0.5	0.458	0.501	0.69
C4	0.446	0.487	0.521	0.377	0.417	0.472	0.665
C5	0.45	0.489	0.509	0.457	0.335	0.5	0.592
C6	0.502	0.544	0.561	0.525	0.51	0.409	0.67
C7	0.529	0.6	0.649	0.597	0.511	0.568	0.583

The $r_{il} + c_{jl}$ and $r_{il} - c_{jl}$ values can be calculated from the total relation matrix (Table 2). The $r_{il} + c_{jl}$ value indicates how important a criterion is, while the $r_{il} - c_{jl}$ value indicates the size of the direct impact of this criterion on other criteria. However, if this value is negative and large, it implies that this criterion is highly influenced by other criteria [4].

The results shown that Aesthetic layout (C1) ($r_{il} - c_{jl}$=0.013), multiple layout style choice (C2) ($r_{il} - c_{jl}$=0.097), storage capacity (C5) ($r_{il} - c_{jl}$=0.205), and system stability (C6) ($r_{il} - c_{jl}$=0.3) were the positively-affected criteria of blog design, it is said, there are "cause" criteria among the seven criteria.

On the other hand, ease of management (C3) ($r_{il} - c_{jl}$=-0.115), ease of registration (C4) ($r_{il} - c_{jl}$=-0.051), and friendliness to beginners (C7) ($r_{il} - c_{jl}$=-0.449) were the negatively-affected criteria, it is said, , there are "effect" criteria.

Therefore, the critical criteria of blog design were found to be "system stability" and "storage capacity", "multiple layout style choice", and "aesthetic layout".

Table 2. The causal influence levels for the criteria

Criteria	$r_{il} + c_{jl}$	$r_{il} - c_{jl}$
Aesthetic Layout (C1)	6.734	0.013
Multiple Layout Style Choice (C2)	7.437	0.097
Ease of Management (C3)	7.401	-0.115
Ease of Registration (C4)	6.819	-0.051
Storage Capacity (C5)	6.458	0.205
System Stability (C6)	7.143	0.3
Friendliness to Beginners (C7)	8.526	-0.449

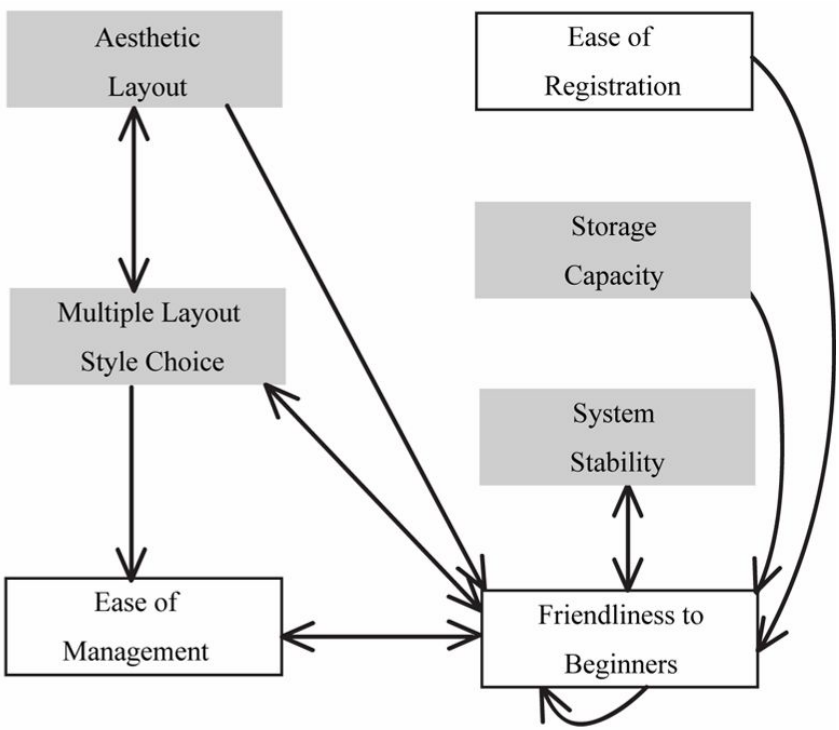


Fig. 1. The impact-relations map for the seven criteria (The arrows indicate the direction of the influence)

3.2 Impact-Relations Map

This research proceeded further by carried thorough discussion with experts where appropriate threshold values were derived in accordance with the identified five factors. The threshold values for design criteria were 0.55. Only values above this threshold were considered, otherwise the relationships would be too complex (Table 1). The threshold values have not standard settings. The thresholds are decided through expert options in this field.

After deciding the threshold values, the DEMATEL impact-relations map could be obtained (Fig. 1). By drawing the impact-relations map for the criteria of each factor, the complicated causal relationships could be visualized. These are summarized below:

Friendliness to beginners (C7) was positively affected by storage capacity (C5) and aesthetic layout design (C1) these two criteria, and formed the interrelatedness with ease of registration (C4), multiple layout style choice (C2), ease of manage (C10), and system stability (C6). Ease of management (C3) was positively affected by system stability (C6) and multiple layout style choice (C2) these two criteria, and formed interrelatedness with friendliness to Beginners (C7). Multiple layout style choice (C2) positively affected ease of management (C3), and formed interrelatedness with friendliness to beginners (C7) and aesthetic layout (C1). Lastly, there was interrelatedness among ease of registration (C4) and friendliness to beginners (C7).

4 Discussion

Based on the results of the empirical study, the researchers make the following suggestions regarding seven design criteria of blog: aesthetic layout, multiple layout style choice, ease of management, ease of registration, storage capacity, system stability, friendliness to beginners.

There are some important messages and implications to be drawn from the impact-relations map (Fig. 1). If a blog developer wants to enhance the design a blog, then system stability, storage capacity, multiple layout style choice, and aesthetic layout are the most influential criteria and should be considered first. In other words, improving these criteria will result in improving the overall blog design criteria.

Although there are numerous excellent studies devoted to blog design, few of these can explain the interaction between evaluation criteria in a systematic way. According to the results of empirical studies, an impact-relations map of the criteria should prove useful in evaluating blog design and visualizing the interrelationships between the criteria. Consequently, the author hopes that this study will make a useful contribution to better understanding blog design.

Acknowledgement. The authors would like to thank anonymous referees for their helpful comments, and Jo-Tzu Chu & Yi-Xian Lee for their assistance in the beginning.

References

1. Guo, Y., Salvendy, G.: Factor structure of content preparation for e-business web sites: Results of a survey of 428 industrial employees in P. R. China. Behavior and Information Technology 28, 73–86 (2009)

2. Nielsen, J., Tahir, M.: Homepage usability: 50 websites deconstructed. New Riders, Indianapolis (2002)
3. Horton, R.P., Buck, T., Waterson, P.E., Clegg, C.W.: Explaining intranet use with the Technology Acceptance Model. Journal of Information Technology 237–249 (2001)
4. Lee, S., Kim, B.G.: Factors affecting the usage of intranet: A confirmatory study. Computers in Human Behavior 25, 191–201 (2009)
5. Guan, S.S., Liu, Y.Y.: Evaluation from the Perspective Of Usability On Wretch Website's Blog Interface Design. Journal of Science and Technology 16, 61–74 (2007)
6. Hsu, C.L., Lin, J.C.C.: Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. Information & Management 45, 65–74 (2008)
7. Tzeng, G.H., Chiang, C.H., Li, C.W.: Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. Expert Systems with Applications 32, 1028–1044 (2007)
8. Chen, J.K., Chen, I.S.: Using a novel conjunctive MCDM approach based on DEMATEL, fuzzy ANP, and TOPSIS as an innovation support system for Taiwanese higher education. Expert Systems with Applications 37, 1981–1990 (2010)
9. Lee, Y.C., Li, M.L., Yen, T.M., Huang, T.H.: Analysis of adopting an integrated decision making trial and evaluation laboratory on a technology acceptance model. Expert Systems with Applications 37, 1745–1754 (2010)
10. Liou, J.J.H., Tzeng, G.H., Chang, H.C.: Airline safety measurement using a hybrid model. Journal of Air Transport Management 13, 243–249 (2007)
11. Wu, W.W.: Choosing knowledge management strategies by using a combined ANP and DEMATEL approach. Expert Systems with Applications 35, 828–835 (2008)
12. Hsu, C.C.: Factors and criteria affecting blog design. Unpublished work. National Chiao Tung University, Taiwan (2010)

Peru Digital: Approaching Interactive Digital Storytelling and Collaborative Interactive Web Design through Digital Ethnography, HCI, and Digital Media

Si Jung Kim and Natalie M. Underberg

University of Central Florida, School of Visual Arts and Design
12461 Research Parkway Suite 500
Orlando, FL 32826-3241
{sjkim,nunderbe}@mail.ucf.edu

Abstract. Digital ethnography is an approach to presenting real-world cultures using the features of digital environments and techniques of narrative. Digital ethnography projects exploit the computational and expressive power of new media to allow audiences to not only learn about, but to also experience something of the culture as well. This approach employs the distinctive features of digital environments such as immersion and interactivity to create new ways to tell cultural stories and enact the research process. This paper presents experiences from a collaborative work where multidisciplinary scholars are involved in creating a cultural website called PeruDigital that presents the culture and history of Peru festivals and related folklore forms for K-12 grade students and individuals interested in Hispanic culture. In addition, this research reflects how digital ethnographers, HCI researchers, and digital media producers are work together in order to create an effective interactive cultural media model.

Keywords: ethnography, cultural media, folklore, participatory design.

1 Introduction

Creating interactive media for cultural learning requires interdisciplinary work between digital ethnography, human-computer interaction (HCI), and digital media to convey the understanding of our real life into a digital domain across various forms such as game, animation, digital storytelling, and so on. Ethnography [1] has been used as an effective way to create interactive cultural media such as a website in which users can learn about culture and history while interacting with the medium [2,3]. For example, Masten and Plowman [4] characterized digital ethnography as the next wave in understanding consumer users living under digital era. Considering digital ethnography's focus on employing digital media as an expressive tool for creative cultural representations, HCI's work on understanding interactions between people and technologies, and digital media's role as content producers, it is important that these three disciplines work together as a design team in order to create effective interactive cultural stories for learning in a cyber world.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 20–28, 2011.

This study aims to explorer that digital ethnography is significant in designing interactive cultural media in that it conveys the understanding of our real life into a digital domain. PeruDigital [5], a digital ethnography project that presents Peruvian festivals on the Internet, is introduced and discussed with regards to how activities are planned to strengthen team cohesion among ethnographers, Andean studies scholars, cultural consultants, HCI researchers, and digital media producers. Lessons learned from the work are presented in this paper, describing how digital ethnographers, HCI researchers, and digital media producers should collaborate to create successful interactive media for cultural.

2 PeruDigital

PeruDigital is a form of online, interactive storytelling that aims to present the culture and history of Peru festivals and related folklore forms for K-12 grade students and individuals interested in Hispanic culture [5]. PeruDigital is an instance of collaborative, interdisciplinary work between digital ethnographers (including folklorists, cultural anthropologists, and Andean studies scholars), HCI researchers, and digital media producers, showing how they collaborated to create an interactive cultural story based on real stories.

A snapshot of PeruDigital is shown in Figure 1, contrasting original material with a digitally converged version. The beneficiaries of PeruDigital are not only individuals who would like to learn about diverse cultures by interacting with a website providing multiple perspectives and situated knowledge, but also multidisciplinary scholars who are associated with collaborative web design which embeds multiple perspectives.

(a) (b)

Fig. 1. A Snapshot of PeruDigital: (a) Original fieldwork photograph (b) Digital version from PeruDigital

The PeruDigital project presents Peruvian festivals and folklore on the Internet through an interactive, immersive, and multilingual Website. The project is based on archive data in the collection of the Institute of Ethnomusicology at the Pontifical Catholic University of Peru-Lima (PUCP), as well as on data collected by the PeruDigital team in Florida and Peru.

The pilot project allows visitors to virtually tour a busy downtown Lima, Peru plaza and the location of the Lord of Agony festival in Afro-Peruvian coastal Piura, Peru. The Lima environment serves to introduce users to the characters, objects, and ideas that relate to the Piura, Peru festival environment (and other future planned virtual festival environments). For example, visitors can peruse information and archive materials in the virtual ethnographer's office, including books, slideshows, video clips, and a fieldwork journal.

The Website presents these archive and ethnographic fieldwork materials from the point of view of scholarship on Peruvian folklore, popular culture, cultural anthropology, and related disciplines. Using an experimental and experiential approach, the Website is designed to embed not just the information that an ethnographer would have but the methodological approach he or she would use to learn that information, as well as the information and insights of those who participate (whether creating, planning, or performing) in the expressive cultural events documented and presented.

As visitors enter the Piura environment, after choosing either English, Spanish, or Quechua, the user encounters an image of the Lord of Agony in Bernal, Piura, the site of the festival that bears the statue's name. The visitor can then navigate to the next scene in order to encounter festival characters and, by interacting with them and the Peruvian ethnomusicology student who provides helpful commentary, learn about the role of these festival figures and what they represent. The Website presents and contextualizes a number of aspects of the festival, including its aesthetics, ethnic and gender identity issues, and history. For example, the visitor can interact with the ethnomusicology student to ask him a question about one of the festival characters, then view a brief archive video of a past performance, and then interact with the festival character himself. By interacting with these characters and the objects they hold (like notebooks or mobile video players), the user can learn about the history of festival stereotypes in relation to questions of identity.

3 Methods

Participatory design [6], which brings together actual users and multidisciplinary scholars, is used in this research. Participatory design is considered to involve multidisciplinary experts and actual users and to see how they collaborate each other. The participatory design team consisted of two groups: 1) Digital ethnographers with cultural consultants who are native Peruvians and 2) A research and technical production group composed mainly of students. The digital ethnographers and consultants group consisted of six individuals, plus a fifteen-person advisory board comprised of international, interdisciplinary scholars. The consults are a group of folk artists, faculty, and graduate students originally from Peru. The research and technical production group consisted of fourteen members that were divided into teams according to their roles. The teams' configurations were six students with a focus on digital media production (including separate roles for graphic design, interactive development, video editing and production, and content asset management); three ethnographic research assistants; and two translators. Later an animator and visual language faculty member joined the team as Art Director to help direct the graphic design aspect, and a HCI researcher to help formalize an interaction model for the project.

A pilot project based on heuristic evaluation [7] was developed over the 2008-2009 academic year (with partnership-building, planning, and content selection completed in 2007-2008), with heuristic evaluation occurring during the 2009-2010 year. The pilot project was developed over the course of monthly team-wide meetings, supplemented by weekly digital media production team and project director meetings. The pilot project was to evaluate the current design of PeruDigital to validate whether the design meets users needs conveying cultural factors through an interactive website. Two teams were formed: an expert advisory board and the UCF students, and each team conducting a task which went through the current design and listed all design problems. This was to measure indirectly whether or not the ethnographers, HCI researcher, and digital media producers worked together well enough to transform the understanding of Peruvian culture into a digital domain. The heuristic evaluation was conducted by email and in-person interviews and solicitation of feedback from the advisory board and cultural consultants, and by a class project in a Digital Imagery class for digital media undergraduates under the direction of Professor Jo Anne Adams.

4 Results and Discussion

The heuristic evaluation revealed a total of 21 design considerations as presented in the Appendix. A card sorting task was conducted for the 21 design considerations to ascertain the characteristics of them and it discovered the four major design consideration: Contents Representation (10, 48%), User Interface (6, 29%), Media Representation (3, 14%), and Extend information (2, 10%). The distribution of the four categories is shown in Figure 2. As the figure shows, the Peru festival and history contents prepared by the ethnographers were not transformed accordingly on the project website by the other researchers. For example, regarding #19 on the 21 design considerations saying highlighting the landscape of the environment, a major cultural asset, the HCI researchers and media producers failed to address emphasizing the visual contents of the website. There are user interface issues and one of them is suggesting adding a drop down box into the navigation links with a sub navigation link so that site visitors can go directly to the information they are interested in. Providing different types of media (e.g., mobile device) and allowing the user to obtain in depth information (e.g., link to relevant web resources) were other issues discovered in the card sorting task.

Another interesting finding was that the visual style of PeruDigital website seemed aimed at children while the text appeared targeted toward adults shown in Figure 3. In the process, the ethnographers transformed archival data based on ethnographic fieldwork into an explorable digital environment. The HCI researchers analyzed the digitized information and identified design requirements in terms of information and interaction design, and then the digital media producers implemented a website according to the design requirements. We realized that there was a gap between the ethnographer and HCI researcher in identifying visual elements to match the text, and this caused another gap between ethnographer and media producers in realizing the visual elements.

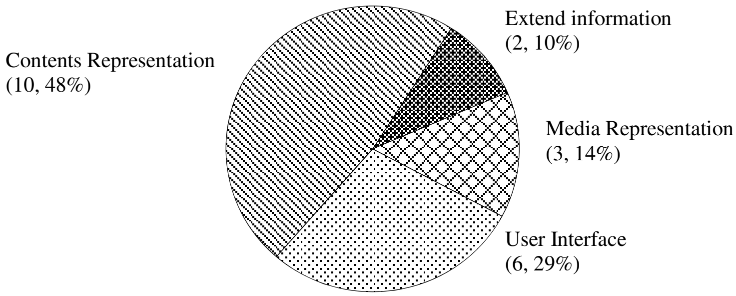
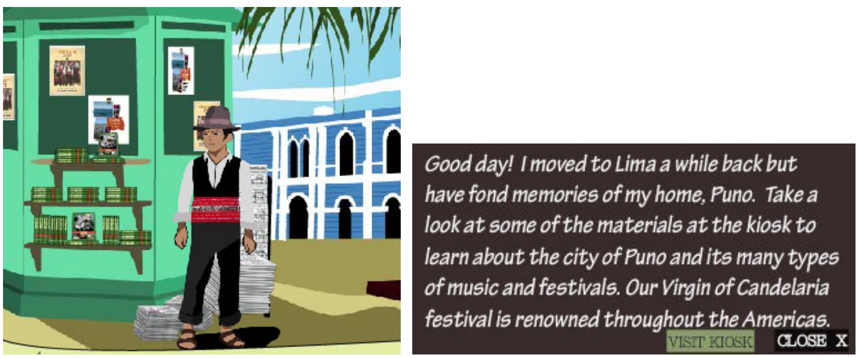


Fig. 2. The distribution of four major design considerations



(a) (b)

Fig. 3. An example of mismatch between visual element and text: (a) Visual element, (b) Text

Figure 4 shows the processes how the three researchers worked together in creating the website. As Figure 4 shows, the interaction process among the three researchers is a sequential flow and we thought that this lead to mismatch between the visual and the text.

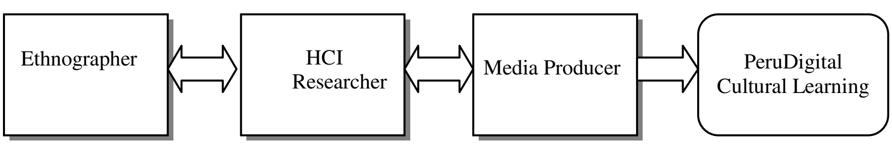


Fig. 4. The Sequential Process when Creating a PeruDigital Cultural Learning Website

Figure 5 is our anticipated interaction process model that was obtained from the previous interaction process represented in Figure 4. In Figure 5, the ethnographer expands his/her interaction with both HCI researchers and media producers in order to address the three design elements: Contents by ethnographer, Information/Interaction design by HCI researcher, and System by media producer. As the new process explains, it is expected that all researchers should interact with each other reciprocally: ethnographer interacts with both HCI researcher and media producer, HCI researcher interacts with both ethnographer and media producer, and media producer interacts with HCI researcher and ethnographer in order to remove any gaps among them.

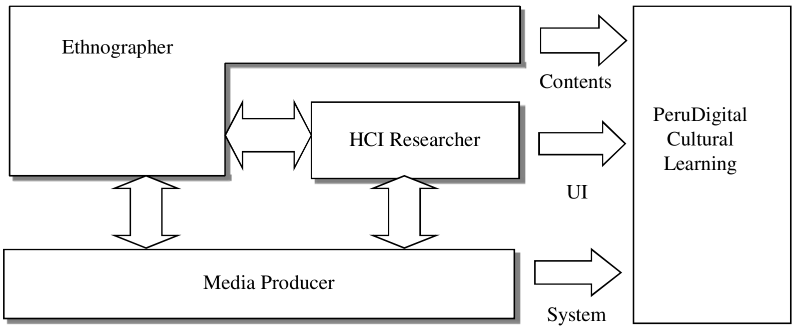


Fig. 5. An Anticipated Interaction Process when Creating Cultural Learning Websites

The project utilized aspects of Participatory Design (PD) in creating the pilot project. The goal was to contribute to the development of a digital media-based visual anthropology and to decolonize knowledge production by using technology to collaboratively represent culture. Watkins (2007) explains: "Phase (1) of the new method comprises a period of due diligence, which informs phase (2) iterative design cycles. These cycles repeat until phase (3), the achievement of desired system/artifact performance" (Watkins 2007: 161). The first phase, "due diligence," includes three stages: completing an organizational observation, making what is called a "domain review," and developing an initial strategy for the project. This was accomplished through two collaboration-building and research trips to Peru to establish collaborative relationships and identify archive materials; completing the research design and establishing a project wiki to facilitate project management and communication; establishing an Advisory Board of scholarly and cultural consultants; and completing initial consultations with cultural consultants. In the prototyping stage, the second step, we conducted Directed Research classes with UCF students to create the framing text for the Website. In addition, we duplicated archive materials for the project, and designed an initial walkthrough of the Website. The downtown Lima environment and the Piura festival space were then completed, to allow us to present a pilot to the Advisory Board and others for evaluation. This project, like others that employ the Participatory Design (PD) model, is an iterative process (involving evaluation and garnering feedback to inform future design cycles).

Although the project utilized Participatory Design particularly in terms of "cultural stakeholders" (cultural consultants, international collaborators), what we learned from our evaluation of the pilot project was to more precisely define the intended user audience earlier in the development process, and build in more evaluative mechanisms to gauge success and identify needed areas of improvement.

5 Conclusion and Future Work

This paper introduced PeruDigital, an interactive storytelling website presenting the culture and history of Peru festivals and related folklore forms. PeruDigital is successful in that it has transformed the understanding of real life Peruvian festivals into digital domains as a website by working together ethnographers, HCI researchers, and digital media producers, however a usability study discovered that it lacked an appropriate match between visuals and text in the pilot project in terms of the perspective of the anticipated audience due to the gaps between the three researchers. This paper also explored how these three researcher groups can work together to remove the gap in creating cultural heritage- and history- based interactive website. An interaction model guiding collaborations among the three researchers for building an interactive website for cultural learning is suggested shown in Figure 4. The model for digital cultural heritage work on the Internet which provides cultural learning through open and guided interactions, while demonstrating how digital ethnography accomplishes collaborative interactive media by bring together multidisciplinary fields such as user interface designers in HCI and web designers in digital media.

The natural extension of this research is to validate the interaction model shown in Figure 6 and conduct another study to ascertain how digital ethnography facilitates the design of effective cultural media with other multidisciplinary researchers.

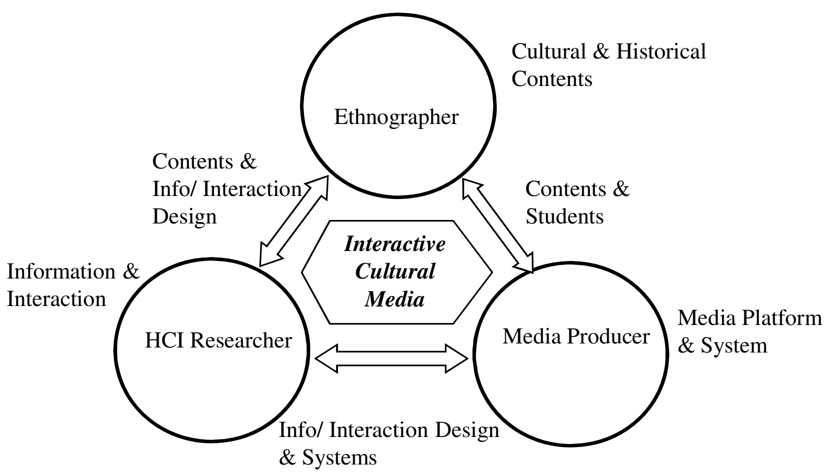


Fig. 6. A Conceptual Representation of Collaborative Interaction Model between Ethnography, HCI, and Digital Media

It is our conjecture that the analysis of interaction behaviors between the three researchers (ethnography, HCI, and digital media) shown in Figure 6 can be translated into a process flow that describes the guidelines, framework, or recommendations for creating effective cultural learning media.

Acknowledgments. The project on which this paper is based received support from a number of organizations at the University of Central Florida (UCF): the School of Visual Arts and Design (formerly Digital Media), Latin American, Caribbean, and Latino Studies program, Center for Humanities and Digital Research, CREATE, Office of International Studies, and the College of Arts and Humanities. In addition, the Pontifical Catholic University of Peru-Lima, in particular the Institute of Ethnomusicology has been a valued collaborator in this research.

References

1. Blomberg, J., Burrell, M., Guest, G. (eds.): An ethnographic approach to design. L. Erlbaum Associates Inc., Hillsdale (2002)
2. Underberg, N.: Virtual and Reciprocal Ethnography on the Internet: The East Mims Oral History Project Web Site. International Journal of American Folklore 119(473), 301–311 (2006)
3. Underberg, N.: Negotiating Puerto Rican Identity in Central Florida and Online. CENTRO: The Journal of the Puerto Rican Studies Association XXII (1), 117–127 (2010)
4. Masten, D.L., Plowman, T.M.P.: Digital ethnography: The next wave in understanding the consumer experience. Design Management Journal 14(2), 75–81 (2003)
5. `http://www.peruvine.org`
6. Schuler, D., Namioka, A.: Participatory Design: Principles and Practices. Lawrence Erlbaum Associates, Hillsdale (1993)
7. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. Paper Presented at the Proc. ACM CHI 1990, Seattle, WA USA (1990)

Appendix

List of critique and suggestions from the heuristic evaluation

1. Do away with the Flash file. It is not universal and takes too long to load.
2. Embed the videos so the viewer has a choice to watch or not.
3. Add a link page, an index listing the links, so the visitor doesn't miss anything.
4. Offer a version of the site with audio of sounds of the city and people speaking their native tongue.
5. Use a more colorful palette like Peruvian Warmth.
6. Use a more recognizable logo and place it throughout the site.
7. Replace the heavy banner on the pages with one that incorporates the logo.
8. Redesign "Peruvine" into a banner with the three languages below it on the Splash page.
9. Place an image of Peru with symbols of Peru on the Splash page.
10. Re-format the page size to 1024 x 768.
11. Identify the buttons in the interactive environment with a glow or border. Use more readily identifiable links through size and placement. (The iPod in the guide's hand is very easy to miss. As a side note, the device will become dated with the next technological version.)

12. Redesign the top navigation bar with greater contrast of color and a typeface change.
13. Instead of using the link “Interactive Environment” on the top navigation bar, perhaps use a drop down menu of places to visit.
14. Perhaps add a drop down box into the navigation links with a sub navigation link so viewers can go directly to the information they are interested in.
15. Replace the arrows with an interactive drop box with places to explore.
16. Adjust the kerning of the titles and text for better design.
17. Be more sensitive to the potential viewer’s cultural heritage. Don’t throw on one Spanish word like “amigo” or “compadre” into an otherwise English text body. The viewer may feel it is condescending.
18. Give the tour guide an identity for greater placement in the story or replace the human with an animal symbolic of Peru.
19. Highlight the landscape of the environment, a major cultural asset.
20. Update the Progress Page often.
21. Use symbols of the festivals as a navigation device. Example: use the diablada masks in the ethnographer’s office to link to the Lord of Agony festival.

Did You Forget Your Password?

Abbas Moallem

San Jose State University
Abbas.moallem@sjsu.edu

Abstract. A quantitative research surveying 390 people with different levels of expertise in computer usage was conducted to understand user behavior from three perspectives: How users make sure that the sites they are using are safe, How users deal with forgotten passwords, How secure is the "security questions". The finding shows users' pattern of behavior in checking security when viewing a web application, the way they deal with numerous passwords and retrieval of the forgotten password by using the security question. The research concludes that most people would be able to answer a variety of security questions for other people in their entourage. Users seem to have significantly different behaviors statistically by age group and level of expertise.

Keywords: Password, Authentication, Security, Reset Password, Password Remembrance, eCommerce.

1 Introduction

Accessing web applications is now part of every individual's daily life. According to internetworldstats.com, nearly 1.9 billion people use the Internet to log into bank accounts or ecommerce to manage their finances or order goods, send and receive emails, or socialize through networking and community sites several times per day.

To access the right location users must point their browser to the right location (the page) and then authenticate themselves with a user ID and password. This process is not as straightforward as it appears.

The first question one might ask is how can users go to a specific web application or website and check if it is really the location that they intended to go to and not a fraudulent one. Users might use different methods to access a specific location. These methods might be typing the URL or domain name in the address bar, using a saved bookmark, or simply clicking a link from another page. The use of any of these methods might require a different effort from the user's perspective, but all also represent risks. For example a slight misspelling when the user enters the domain name into the address bar as well as a misleading link might take the user to a fraudulent site and easily acquire the user's authentication information (Keats, 2007, Edlerman, 2008).

Fraudulently acquiring people's sensitive information such as user ID, passwords, and credit card details through a deceitful link usually by e-mail or instant messaging (known as phishing) is widely known and documented security issue. However, many people still might click on the deceptive link and enter their private information. Researchers report a

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 29–39, 2011.

variety of techniques to use web history to access users' information (Naone, 2010, Jagatic et al, 2005, CIO Council, 2009, Castelluccia et al, 2010).

The second question is how would users check to see if the site is using a safe connection or is not a phishing site or parking page? To check the validity of a site, a variety of methods were offered, in this case the most common techniques for validating the site, by the order of best security offered are: checking external certificates, checking connection HTTPS, checking site reputation by looking at the appearance and content of the site or rely on third-party applications to check the validity of the visited site.

Among all of the above checking an external certificate seemed to be the most reliable method. Although this method seems to be the most reliable checking method, it is still not 100 percent secure since the hacker can for example go to the user's computer and change the list of certificate authorities to make it trust an un-trusted certificate authority.

Checking the HTTPS connection, which is also related to the certificate validation, is another reliable method to check the validity of a web application site. The main idea of HTTPS is to create a secure channel over an insecure network. An HTTPS connection uses certificates to validate trusted authorities. If that check fails, then the browser will warn users. However some research suggests that site-authentication images are ineffective and the users do not withheld their passwords even when the site is not using the HTTPS connection (Schechter et al. 2007).

Another method to check the validity of a site is by paying attention to the look and content of the site but this method alone would not be a reliable security check. Most of the phishing sites simulate a total visual replica of the original site.

A further technique to check the security of the site visited is to rely on third-party applications. These applications check the validity of the visited site and inform the user about the site. Despite their use of safety features, they require the users to have a basic understanding of Internet security features and the risks and notifications or warnings offered by these applications.

The next step is for the user to pass the authentication phase. This process consists of authenticating users through the use of one or more of three "factors" (FFIEC, 2008): something the user knows (e.g., password), something the user has (e.g., smart card); or something the user is (e.g., fingerprint).

Among these factors the authentication through user ID (or user name) and password seems to be the most common one.

With accumulations of the web application accounts, users' behavior was the subject of a variety of studies that include: the diversity of the password format requirements, too-easy-to-guess passwords and methods to prevent dictionary attacks, number of password that each user uses to access different accounts, password complexity, difficulty to remember passwords and user names for different account, self-resetting password, and password retrieval. (Florencio & Herley, 2007, Gaw & Felten , 2006, Englert & Shah, 2009, Forget & Biddle, 2008, Karaca & Levi, 2008, Weir et al, 2010).

According to research cited above, users do not seem to be very conscious about the security issues or do not take into consideration the reliable cues. Users use a limited number of passwords that are sometimes very simple if they are not forced by password policies to make their password more complicated. The users may have many web accounts but that does not mean they have different passwords for each account.

To improve security, web applications now use several features to address forgotten passwords and ID reset. These functionalities consist of providing a user with a self-reset password and user ID capability requiring them to answer several security questions, and then, if the answers are correct, the passwords are sent through email to the email address on file. Then the user must use that temporary link or password to login and change their password.

Despite the various studies reviewed above, there seem to have been no studies that have been conducted from the perspective this study uses.

Consequently this research tries to answer the following questions:

- How conscious are users about the security when they login to a web application?
- How do they check if the site is secure?
- How often do they use password or user name reset features?
- How much security is offered by security questions in these reset features?

We have considered two parameters that might be distinctive among the users: age and expertise. We hypothesized that younger and more expert users might have a different behavior in all the above factors.

2 Methods

An online survey was designed containing three groups of questions: questions regarding respondent profile, security, and forgotten password reset features and security questions.

A message, with a link to online surveys, then was sent to LinkedIn groups, investigator's former students, and posted online in several groups to voluntary participants to take the survey. A total of 390 responses were collected. All data was collected from mid- March, to mid- May 2010. The data collected through this survey was then analyzed using statistical analysis.

Participants. Overall 64% of the respondents were male and 36% female. The percentage of the respondents by age group is: 41% over 55, 21% 46-55, 32%, 36-45, 25%, 26-35 and 9%, 18-25. 57% of participants have a graduate school level of formal education and 34% a college degree. The participants were asked to self-evaluate their level of expertise in computer usage. 34% evaluated their expertise as Expert, 45% as Advanced, 20% as Intermediate.

3 Results

In this section we will review responses collected for each group of questions: Security, Forgot password and Security Questions.

3.1 Security Questions

To understand the user behavior in checking the security of a site, 3 questions were asked. In the following section, we will analyze the answers provided to these questions.

Q-S1: How do you find the site you intend to log into?
Overall 62% of the responses indicate that participants type the URL address in the address bar, 59% use their bookmarks, and 36% use search. Typing the URL into the address bar seems to be favored more by "expert" users. However, older participants seem to favor using bookmarks 72% among the over 55 age and 73% among age 46-55 but 52% among the 26-35 age group.

Q-S2: Once you get to the site that you have requested, how do you make sure that the displayed site is a real site and not a fake or fraudulent site?
Overall the look of the application is the primarily checked option (51%) followed by checking browser certificate validation (43%). It is also interesting to observe that 13% of the participants declare not being aware of any of the suggested techniques for checking the validity of the site. Only 18% of the respondents use a third-party application to check the validity of the site.

Looking at this data shows that checking the browser certificate validation seems to be used more among the expert users (62%). Overall it seems that the main way to check validity of the web application site is the look of the site even among the expert users (41%). Interestingly, 12% of self-evaluated advanced user and 7% of expert users declare not being aware of any method to check the validity of the site.

The data analysis by age group reveals that the age group of 36-45 and 46-55 use more advanced validation methods such as external or browser certificate validation. Overall users seem to rely more on the look than on the use of advanced validation techniques such as certificate valuation.

Q-S3: Whenever you are entering your personal information on an Internet site such as a bank ecommerce or registration page, how do you know if the site is secure and that you have a secure connection?
The participants could select one or more options in the list. The results show that considering all participant answers 64% say they rely on HTTPS, 44% SSL lock, 43% say they use only sites with known reputations or they rely on VeriSign or similar logos (22%).

Looking at the data by self-evaluated level of expertise reveals that reliance on HTTPS and domain name raises by user's level of expertise. For example, fewer intermediate users (27%) check HTTPS versus expert users (82%) or check domain name (63%) among the expert users versus (30%) intermediate users.

3.2 Forgot Password

To understand the user behavior in password retrieval, seven questions were asked of the participants. In the following section, we are analyzing the answers provided given to these questions.

Q-P1: Overall how many different passwords do you use to log in to different web applications or websites?

The results indicate that a negligible number of people use one password. 7% use only two passwords but 49% of the respondents use three to five passwords. Users belonging to the self-evaluated "expert level" seem to use more passwords 48% of use more than 5 passwords. It seems that people with a lower level of expertise use fewer passwords than users with a higher level of expertise. Analyzing the results by age group reveals that older users use more than 5 passwords (over 55 age, 54% and age 26-36 42%). Considering all results, it seems that almost the same ratio of people (33% to 57%) across the expertise and age groups use three to five passwords.

Q-P2: How often do you forget your password to log into an application such as a bank account that you use often and for which your user ID and password are not already saved by the browsers?

22% declare they never forget their password for the frequently used application, 51% seldom, and 23% sometimes.

Q-P3: How often do you forget your password to login to an application that you use rarely and your user name and password are not already saved?
The results show that 35% of the respondents often forget their password and 39% sometimes forget. Interestingly 4% of respondents never forget their password. It seems that older users forget their password more often. 32% of people among age group 26-35 forget their password sometime versus 50% among 46-55 age group. However, the decline in the forget rate for the age group of over 55 is 31%, which might indicate that users keep a record of their passwords.

Q-P4: How often do you use the forgot password feature to reset or find your password or user ID/name for applications you use rarely?
The results show 64% of participants seldom use the reset password feature versus 5- to 6% percent every time or often.

Q-P5: Which of the following methods do you like the best to find your user name or password?
The results show that 41% of the participants prefer to answer security questions and receive their password or reset link by email. 26% prefer to answer security questions and then immediately reset the password, 32% prefer to click forgot password/user ID, enter and receive reset password link by email, and almost nobody wants to call support for resetting their password. However, preference for clicking forgot password/user ID, enter and receive reset password link by email is stronger among younger age groups and experts (26-36, 38%, 36-46, 37%, and expert group 42%).

Q-P6: How often do you customize the security question to find your password?
The results shows 25% often customize the question, 28% never do, and 45% sometimes do. The percentage is higher among the more advanced users quite often, 33%, sometimes 39%, but not necessarily different among the age groups.

Q-P7: Do you use any 3rd party application or hardware to store your password and login information?
Overall 70% do not use any third-party software is higher among more expert users (Expert group 48%) but seems not much different among the age groups.

3.3 Security Questions

The participants were asked to estimate the number of people for whom they might know the answers to the questions that are generally asked during user ID and password reset. (Table 1)

Table 1. Security Questions List

For how many people can you answer the following questions?	
In what city was he/she born?	What was his/her first job title?
What is the name of his/her first pet?	What brand was his/her first car?
What is the name of the first company he/she worked for?	What is the name of his/her pet?
What is the name of his/her mother's maiden name?	What is his/her favorite teacher?
What is the name of his/her paternal grandmother?	What color was his/her first car?
What is the name of his/her closest friend at high school?	What color is his/her favorite color?
What was his/her favorite subject in high school?	
What is the name of his/her childhood best friend?	

The results show that for most of these common password reset questions participants know the answer for one or many persons. The more general questions such as in “What city you were born?” only a negligible fraction of people (1%) did not know anybody else’s place of birth. Fewer people knew the answers to questions that required older memory information such as, “your favorite teacher”. 51% did not know anybody’s information about favorite teacher. Even for this question 49% knew at least information about one person. (Figure 1)

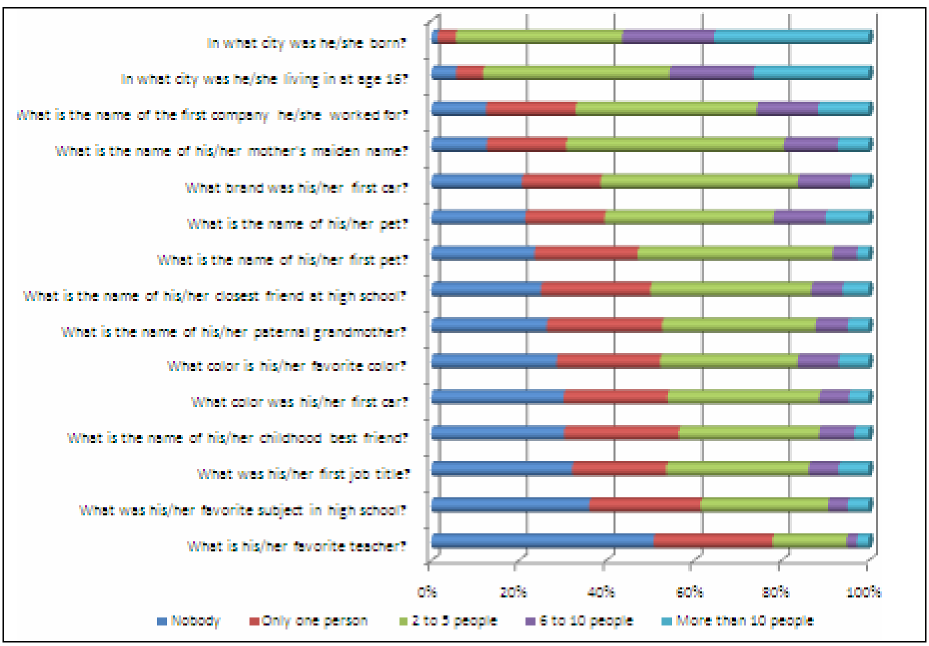


Fig. 1. For how many people can you answer the following questions?

4 Analysis

Statistical data analysis using ANOVA single factor was performed to identify whether there is any significant differences among the groups of participants based on age or level of expertise to verify the following hypotheses.

Participants significantly differ on how they check the security of a web application based on their Age group and level of expertise.

The ANOVA test results indicate no significant differences between the different age groups for questions S1, S2, P3, P6, and P7. The ANOVA reveals significant differences for questions P1, P2, P4 and 5. P-values are 0.69. 0.60 0.75 and 0.53 (Table 2).

The ANOVA test results indicate no significant differences between the different levels of expertise levels for questions S1, S2, S3 and P2, P3, P5 and P6. The ANOVA reveals significant differences for questions P1, P4 and P7. P-values are 0.58., 0.65, and 0.58.

A significant difference was also noticeable between and within age groups for Question P5 and for Question P7 between and within groups based on level of expertise. (Table 3)

This result tends to indicate that the subjects have different behavior in the number of passwords based on different age groups and level of expertise. This seems to be the same for the number of passwords and how often they use the forget password for web applications they rarely use.

Table 2. ANOVAs Single Factor Test Result –Age Groups: 26-35, 36-45, 46-55 & Over 55

Question	Source of Variation	SS	df	MS	F	P-value	F crit
P1	Between Groups	600.4	3	200.13	0.49	**0.69**	3.23
	Within Groups	6490.8	16	405.67			
	Total	7091.2	19				
P2	Between Groups	600.4	3	200.13	0.63	**0.60**	3.2
	Within Groups	5044.8	16	315.03			
	Total	5645.2	19				
P4	Between Groups	600.4	13	200.13	0.40	**0.75**	3.2
	Within Groups	7914.8	16	494.75			
	Total	8515.2	19				
P5	Between Groups	600.4	3	200.13	0.76	**0.53**	3.2
	Within Groups	4190.8	16	262.92			
	Total	4791.2	19				

The participants seem to prefer typing a URL or using a bookmark to navigate to the desired web application site. The results also suggest that advanced and expert users tend to prefer typing the URL whereas older users tend to prefer bookmarks. This might be related to ease of typing and movements as younger users might have more rapid text entry skills than older adults. Research associates aging with changes in motor skills, including slower response times, decline in ability to maintain continuous movements, distribution in coordination, loss of flexibility, and greater variability in movements (Rogers & Fisk, 2000).

Table 3. ANOVAs Single Factor Test Result - Level of Expertise: Intermediate, Advanced, & Experts

Question	Source of Variation	SS	df	MS	F	P-value	F crit
P1	Between Groups	950.53	2	475.26	0.55	**0.58**	3.88
	Within Groups	10234.4	12	852.86			
	Total	11184.93	14				
P4	Between Groups	650.53	2	475.26	0.44	**0.65**	3.88
	Within Groups	12770.4	12	1064.2			
	Total	13720.93	14				
P7	Between Groups	2376.33	2	1188.16	0.64	**0.58**	3.88
	Within Groups	5505	3	1835			
	Total	7881.33	5				

The results of how participants ensure the site is a real site and they are not looking at a fake site suggests that no matter the age or level of expertise the primary verification seems to be the visual aspect of the site followed by browser certificates check. More advanced users tend to rely on external validation certificate check. The use of third-party validation is less than 20%. However according the New York Times (Helft M. 2010), the "number of such third-party "certificate authorities" has proliferated into hundreds spread across the world, it has become increasingly difficult to trust that those who issue the certificates are not misusing them to eavesdrop on the activities of Internet users, the security experts say".

The other most alarming factor is that 13% of participants declare they do not know any method to check the validity of a site. This number indicates the potential risk of these users to become victims of all types of fraud. The participants' comments indicate that some users experienced an invalid certificate for a valid site. This behavior makes users not really trust the browser validation test as one participant's comment indicated, "Certificates have proven to be worthless. Trusted sites have invalid certificates (even Microsoft at times!)." Many participants indicated visual verification of the URL as the primary check of the site validity. Or "according to the Electronic Frontier Foundation, more than 650 organizations can issue certificates that will be accepted by Microsoft's Internet Explorer and Mozilla's Firefox, the two most popular Web browsers. Some of these organizations are in countries like Russia and China, which are suspected of engaging in widespread surveillance of their citizens" (Helft M. 2010). In the Jones et al. (2007) study, it was also found that most respondents were unfamiliar with authentication technologies and suggested that though users are typically concerned about privacy and security, they do not necessarily understand how these issues are impacted by the use of digital identities.

On how a user knows if the site is secure and that they have a secure connection, the results are even more alarming. 6% of all participants indicated that they do not check anything at all. However, we observed a surprising 64% of users check to see if the site uses HTTPS 45% check the domain, 22% pay attention to third party validation such as VersiSign, 44% check the browser lock and 28% pay attention to security image. It seems that the younger group checks security less than the older group. Checking the security, according to Herley (2009), seems to have a direct relationship with the cost of an attack and greater indirect cost of effort by users.

These results are also consistent with other research on security indicating that users did not notice the change in validity of page images (Schechter et al. 2007)

On the remembrance of passwords and password retrieval, the results indicate that a negligible number of people use one password. 7% use only two passwords but 49% of the respondents use three to five passwords. Users belonging to the self-evaluated "expert level" seem to use more passwords 48% use more than 5 passwords. It seems that people with a lower level of expertise use fewer passwords than users with a higher level of expertise. Analyzing the results by age group reveals that older users use more than 5 passwords (over 55 age, 54% and age 26-36 42%). Considering all results, it seems that almost the same ratio of people (33% to 57%) across the expertise and age group use three to five passwords.

The low number of passwords generally used by users indicates that most users use the same category of passwords for several applications. Interestingly, the lower number of password retrieval among the older users group indicates that many users keep a written record of their password on paper. Writing down passwords has been reported by previous researchers (DeAlvare, 1998, Adams & Sasse 1999). DeAlvare reports that 50% of questionnaire respondents wrote their passwords down in one form or another.

An important percentage of the participants 22% declare they rarely use the passport retrieval for frequently used accounts and 27% seldom use it for not frequently used account. This might be another indication that users keep a written version of their password, or use the same password for most of their accounts. The higher percentage of users who rarely or seldom use re-set password, among the older age group, is supporting data and an indication of this behavior.

The findings of this study support the previous studies that a user who has more web applications does not necessary have more passwords. Consequently, it can be extrapolated that the users use the same password over and over for the types of accounts they use. Different password requirements for each application seems to be the main user issue since that would require users use the reset password. When applications require users to change their password frequently or will not allow the user use previously used passwords, this creates a huge usability issue for users.

All things considered, users still frequently use re-set password features and in this case the results support common sense that users do not like to contact customer support to reset a password. Around 25-29% of all ages and levels of expertise prefer to be able to re-set the password immediately with opening an email client to view the link or new temporary password. This percentage is higher (36%) among the over-55 age group.

Despite the fact that 41% of users like to answer the security questions and re-set their password (almost the same ratio among age groups and level of expertise), 28% percent of participants (26% to 30% among all age groups and level of expertise) declare that they never customize the security questions to find their password.

Participants seem neither to know nor to trust third-party software applications to manage their password. It seems more expert level users use a third-party application more often to manage their passwords. Considering that some password management applications seem themselves to be very insecure or users with multiple browses can't take their password from one computer to another, it seems third-party software applications cannot resolve user password issues.

Consequently, answering the security questions seems to be offered often by application to retrieve the password or user ID/name. The common feature is answering a variety of preset questions. In the study, we were aiming to find out how many people were able to guess the answers to the security questions for someone else. This study shows that out of 15 typical security questions, even in the best case only 50% of people do not know the answers for somebody else in their entourage. Only the security questions that come more from the participant's episodic memory such as their favorite teacher or favorite subject in high school perform better. Consequently, there is a strong probability that somebody in a close one-person circle might get access to somebody else's accounts through family disputes, divorce cases, small partnership, and so on.

These results indicate that for Q-P1 "Overall how many different passwords do you use to log in to different web applications or websites?" the participants show a significant difference between and within groups based on both age and level of expertise.

5 Conclusions

The finding of this study supports previous studies in identifying the number of passwords and the security and degree of users' awareness, and concludes that most users are not aware of any way to check the security when viewing a web application. Many users use password retrieval by using the security question. However, most people would be able to answer a variety of security questions for other people in their entourage. People have a tendency to use a very small number of passwords, and they often keep track of those passwords most likely in paper format since a very strong percentage of people respond that they rarely use passport retrieval, especially among the older age group. Users know answers to the security questions of several people around them, and they might be able to answer them and get access to the password if they can get control of the email box.

Users seem to have significantly different behaviors in the number of different passwords they use to log in to different web applications or websites based of their age group and level of expertise. They also statistically differ by age group and level of expertise in the "reset password" feature frequency of use. Users are also differentiated by level of expertise in using 3rd party application or hardware to store password and login information. They are also set apart by age group in the methods used to find their user name or password.

This study needs to be expanded to a more diverse population so that more details of the behavior may be investigated through qualitative evaluation or each subgroup. The users' behavior in this field also evolves based on the user interface offered, but different web applications affect this tendency. Further research is needed to understand how users' behavior changes and to understand their needs and issues.

Acknowledgments. I would like to thank Mr. Mark Schertler, Security Architect at Axway, Inc. for his contributions to this study, Alice and Kenneth Kemp for editing the text, Shabnam and Nassim Moallem for their assistance in data collection and data exportation of this research.

References

[1] Adams, Sasse: Users Are Not the Enemy. Communications of the ACM(1999)

[2] Castelluccia, C.1., De Cristofaro, E., Perito, D.: Private Information Disclosure from Web Searches (The case of Google Web History) (2010), http://planete.inrialpes.fr/~ccastel/PAPERS/historio.pdf (accessed on September 23, 2010)

[3] CIO Council: Guidelines for Secure Use of Social Media by Federal Departments and Agencies, http://www.cio.gov/Documents/Guidelines_for_Secure_Use_Social_Media_v01-0.pdf (accessed on October 21, 2010)

[4] DeAlvare, A.M.: A framework for password selection. In: Proceedings of Unix Security Workshop II, Portland (August 29-30, 1998)

[5] Englert, B., Shah, P.: On the Design and Implementation of a secure Online Password Vault. In: ICHIT 2009, Daejeon, Korea, August 27-29 (2009)

[6] FFIEC, Federal Financial Institutions Examination Council: Authentication in an Internet Banking Environment (2005), accessed on http://www.ffiec.gov/pdf/authentication_guidance.pdf

[7] Forget, A., Biddle, R.: Memorability of Persuasive Passwords. In: CHI 2008 Proceedings, Florence, Italy. ACM, New York (2008) 978-1-60558-012-8/08/2005

[8] Gaw, S., Felten, E.W.: Password Management Strategies for Online Accounts. In: Symposium on Usable Privacy & Security (SOUPS), Pittsburgh, PA, USA, July 12-14 (2006)

[9] Helft, M.: Experts Warn of a Weak Link in the Security of Web Sites New York Times Published on (August 13, 2010), http://www.nytimes.com/2010/08/14/technology/14encrypt.html?_r=1 (accessed on August 23, 2010)

[10] Herley, A.: So Long, And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users, Microsoft Research (2010)

[11] Jagatic, T., Johnson, N., et al.: Social Phishing. ACM, New York (2005), http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf (accessed on October 21, 2010)

[12] Jones, L.A., Antón, A.I., Earp, J.B.: Towards Understanding User Perceptions of Authentication Technologies. In: WPES 2007, Virginia, USA, October 29 (2007)

[13] Karaca, K., Levi, A.: Towards a Framework for Security Analysis of Multiple Password Schemes. In: EUROSEC 2008, Glasgow, Scotland, March 31 (2008)

[14] Keats, S.: Cashing in on Typos (2007), http://www.mcafee.com/us/security_insights/archived/oct_2008/si_oct5_08.html (accessed on May 19, 2010)

[15] Keats, S.: What's In A Name: The State of Typo-Squatting (2007), http://www.siteadvisor.com/studies/typo_squatters_nov2007.html (accessed on 05/19/2010)

[16] Naone E: Peeking Into Users' Web History, Technology Review (April 21, 2010), http://www.technologyreview.com/web/25159/?a=f

[17] Rogers, W.A., Fisk, A.D.: Human Factors, applied cognition and aging. In: Crailk, E.I.M., Salthouse, T.A. (eds.) The Handbook of Aging and Cognition. Lawrence Erlbaum Associates, Mahwah (2000)

[18] Schechte, S.E., Dhamija, R., Ozment, A., Fischer, I.: The Emperor's New Security Indicators. An evaluation of website authentication and the effect of role playing on usability studies. In: The 2007 IEEE Symposium on Security and Privacy, Oakland, California, May 20-23 (2007), accessed on http://usablesecurity.org/emperor

The Layout for the User-Friendly Manual: Case Study on an Internet Set-Up Manual

Momoko Nakatani[1,2], Takehiko Ohno[1], Yurika Katagiri[1], Ai Nakane[1], and Shuji Hashimoto[1,2]

[1] NTT Cyber Solutions Laboratories, Human Interaction Project, 1-1 Hikari-no-oka, Yokosuka-shi, Japan
[2] Waseda University, Faculty of Science Engineering
{nakatani.momoko,ohno.takehiko,katagiri.yurika, nakane.ai}@lab.ntt.co.jp

Abstract. We propose two design concepts for the user-friendly manual and compare them in an experiment. The first concept, which focuses on user comprehension, is to use one picture of the completed wiring. The second concept is to uses a series of steps from left to right with the goal of making the user follow the order. Trials show that participants presented with material based on the first concept tend to follow their own mental-model rather than the manual. Material based on the second concept also failed to make users follow the order. Some implications for the refinement of manual design are derived based on the results.

Keywords: Usability, manuals, documentation, technical communication.

1 Introduction

Developing an effective design concept for the user-friendly manual is crucial given the extremely large, are growing, number of home appliances. Consumers are well known to seldom read or follow the instructions in the manner intended by the designer, resulting in a lot of frustration [1],[2].

While software configuration can be automated, users still have to manually perform the wiring task. As the home-network is now entering the home, the action of wiring the devices has become more complicated, and the importance of developing user-friendly manuals for the wiring task is increasing. Service providers often provide a one-sheet manual that places all crucial information on a single piece of paper to moderate the users' resistance to reading. Our challenge is to construct a method for designing user-friendly lucid one-sheet-paper manuals for the reluctant user.

In this paper, we propose two design concepts for wiring manuals. The concepts and details of the experiment are described in the following section. Results are then introduced and discussed.

2 Design Concepts

As the test case, this paper, considers the replacement of a home router; its connection diagram is shown in Fig.1. The user is expected to replace the old router, which has

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 40–45, 2011.

failed, with the new router. The procedures for unplugging and reconnection are shown in Fig.2.

Many of the wiring tasks of recent home appliances contain steps that must be done in a sequence. In the test case, the user must “(5) Take off the stand” before “(6) Opening the lid” to “(7) Unplug the optical fiber”. Step “(11) Plug the optical fiber in” cannot be done before step “(9) Take the cap off the connector”.

We devised two design concepts: “Duplicate Layout (DL)” and “Ordered Layout (OL)”. The concept of the DL is to make users understand the overall wiring layout.

Unplugging Procedure	Plugging Procedure
(1) Unplug power cord from the wall	(8) Open the lid of the new router
(2) Unplug optical fiber from the wall	(9) Take the cap off the connector
(3) Unplug LAN cable from the old router	(10) Move the part
(4) Unplug telephone line from the old router	(11) Plug the optical fiber
(5) Take off the stand of the old router	(12) Close the lid of the new router
(6) Open the lid from the old router	(13) Plug the optical fiber into the wall
(7) Unplug the optical fiber	(14) Plug the lan cable into the new router
	(15) Plug the telephone line into the new router
	(16) Plug the power cord into the new router
	(17) Plug the power cord to the wall

Fig. 1. This shows the procedures for replacing a broken router. (1)-(7) is the procedure for unplugging the cables from the old router and (8)-(15) is for plugging the cables into the new router.

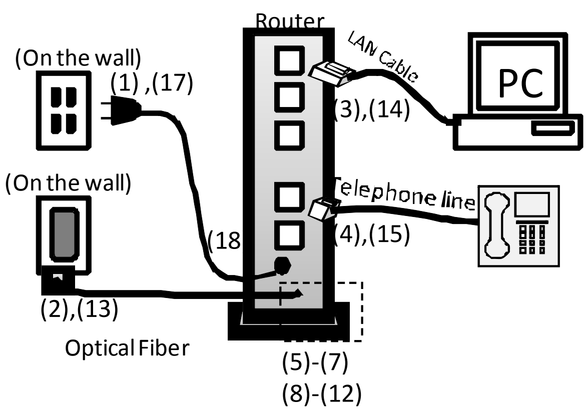


Fig. 2. This shows the wiring diagram. (1)-(15) that corresponds to the unplugging and unplugging procedures which is described in Fig.1.

Pictures of two devices that should be connected or disconnected should be placed next to each other so that the users can easily understand what devices should be connected or disconnected. The OL, on the other hand, places the procedures in order to encourage the users to follow the order.

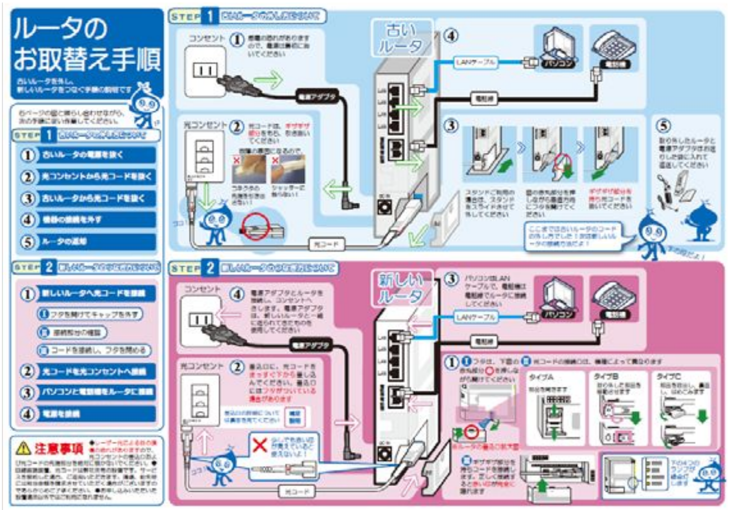


Fig. 3. Duplicate Layout manual (DL). The spatial locations of the upper components are copied by the lower components.

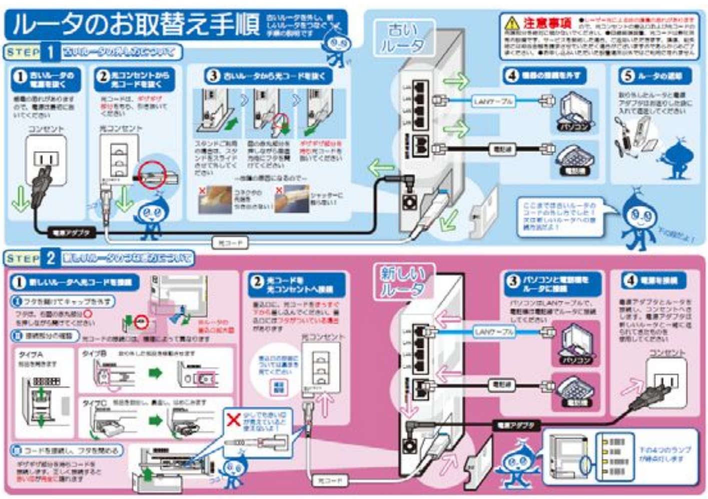


Fig. 4. Ordered Layout manual (OL) where both unplugging and reconnection actions run from left to right. However, the lower pane reverses the spatial locations of components.

We used these design concepts to design the two manuals shown in Fig. 3 and 4. The upper pane explains the unplugging steps and the lower pane the reconnection steps. For example, the power connector occupies the top left corner in both manuals. In the DL (Fig.3), the pictures of the outlet on the wall, PC, and the telephone are placed next to the router. The DL also copies the spatial locations of the components so that user can understand the overall wiring layout. However, reconnection actions run from right to left. OL (Fig.4) reverses the spatial locations of components, and the picture of the outlet is not placed next to the picture of the router, but both unplugging and reconnection actions run from left to right.

3 Experiment

3.1 Method

To find out how users actually execute the procedures after being provided with these two manuals, we conducted an experiment. We hired 15 novice Internet users (30-50 years old) and divided them into two groups, DL (7 subjects who used the Duplicate Layout manual) and OL (8 subjects who used the Ordered Layout manual. Each subject was told to replace the broken router.

3.2 Results

(1) Success rate
85.7% (6 out of 7) of DL subjects, and 87.5% (7 out of 8) of OL subjects succeeded in replacing the device without anybody's help, and there were no significant differences between the two groups.

(2) Order of the action
Average number of times that users went back one or more steps is shown in Table 1(a). If the user performed step(2) and then step(1), for example, the number was incremented. DL subjects demonstrated more step reverses than OL subjects, but the difference was not significant ($p>0.05$).

Average number of times they switched between unplugging and reconnection actions is shown in Table1(b). If the users performed all the procedures in order, the number was counted as "1"; reconnection follows unplugging. The DL subject with the highest switch number first unplugged the LAN cable from the old router, and then plugged it into the new router, and then unplugged the telephone line from the old router, and plugged it into the new router. Overall, DL subjects tended to move between the old/new router more often than OL subjects.

Table 1. Number of the times the participants reversed a step and switched between unplugging and reconnection actions

Heading level	Condition	Average	Min	Max
(a)Number of times they reversed a step	DL	4.57(SD=2.22)	1	8
	OL	3.25(SD=1.28)	1	5
(b)Number of times they switched between unplugging and reconnection actions	DL	5.57(SD=3.95)	1	12
	OL	2.86(SD=1.84)	1	5

(3) Order of the action in "strictly ordered procedures"
The test case included two strictly ordered procedures. First, step "(7) Unplug the optical fiber" cannot be done unless step "(5) Take off the stand of the old router", and step "(6) Open the lid of the old router" have been performed. Second, step "(11) Plug the optical fiber in" cannot be done before step "(9) Take the cap off the connector".

Most DL participants didn't follow the order required; 3 out of 8 participants didn't order steps (6) and (7) correctly, and 4 out of 8 misordered steps (9) and (11). OL participants were guilty of the same errors; 2 out of 7 participants misordered steps (6) and(7) and 6 participants misordered steps (9) and (11). These participants wasted time and energy in order to get the order right.

4 Discussion

With reference to the results in Table1(b), the DL switched frequently between the old router and the new router; unplug this and immediately reconnect it to the new device. It is known that constructing appropriate mental model leads users to execute the procedures faster by simplifying inefficient procedures [3][4]. So we assume that participants who used DL constructed their mental-model and relied on it rather than following the procedures written in the manual. If the procedures are not required to be followed in order, this DL approach is efficient for the faster execution and the flexible error recovery.

However, if the order is strictly defined, the user may encounter some trouble. DL participants actually didn't follow the strictly ordered procedures (steps (6),(7) and steps (9), (11)), which led to extra effort to go over the procedure again.

One-page manuals should, therefore, be designed to clearly indicate step sequencing requirements. If the order is inflexible, the manual layout should focus the user on one step at a time. Otherwise, layouts that enhance the user's understanding of the whole picture are suitable [4].

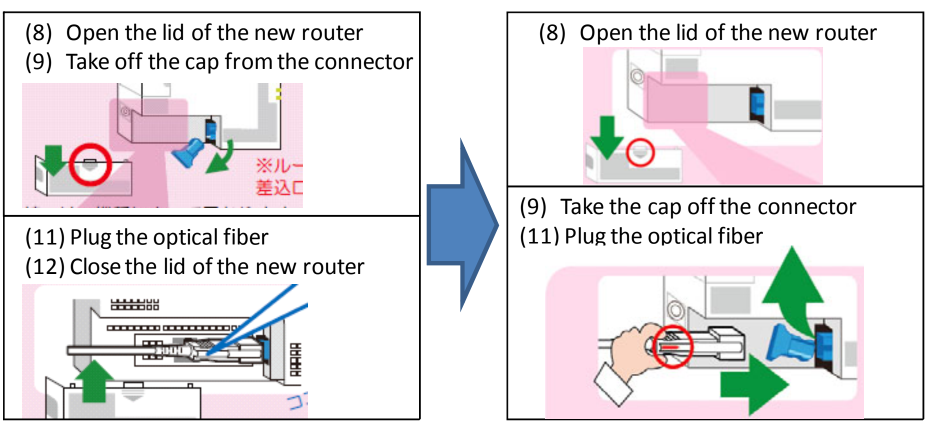


Fig. 5. Left side pane shows part of the original manual: steps (8)-(11) (steps (8) and (9) are shown in one picture). Right side pane shows the refined version: steps (9) and (11) are combined.

While the design concept of the OL was intended to enforce the step order, it was not really effective in doing so as shown in Table1(a). Our solution is to elaborate the design in two ways. First is to eliminate redundant explanations (e.g. title of each step, names of each cable) so as to emphasize the order information (step number). The second idea is to change the way of grouping the pictures to emphasize the inflexible step sequences. Fig.5 shows an example of a refined manual for steps (9)-(11). The left pane shows the manual used in the experiment; steps (8) and (9) were combined in one picture and step(11) was presented separately. However, combining steps (9) and (11), in the right pane, is expected to ensure that the user follows the order.

We assume that this enhanced design will improve the efficiency of users' actions. We intend to validate the effectiveness of these design refinements.

5 Conclusion

We have described two design concepts to develop the user-friendly manual. By comparing two design layouts, the DL which focuses on helping the user to understand the overall wiring arrangement, tends to let the user follow his/her own mental-model rather than the manual. Past research suggests that if the order of steps is flexible, the manual layout that enhances the user's understanding of the overall arrangement is most suitable. On the other hand, if step order is inflexible, the manual layout should focus the user on performing one step at a time in the correct order.

Because the original OL was not as effective as hoped, we proposed two enhancements and intend to confirm their efficiency in subsequent research.

References

1. Carroll, J.M., Rosson, M.B.: Paradox of the active user. In: Carroll, J.M. (ed.) Interfacing Thought: Cognitive Aspects of Human-Computer Interaction, pp. 80–111. MIT press, Cambridge (1987)
2. David, N. G., Karen, W.: Why don't people read the manual? In: SIGDOC 2006: Proceedings of the 24th Annual ACM International Conference on Design of Communication, pp. 11–18 (2006)
3. David, E.K., Bovair, S.: The Role of a Mental Model in Learning to operate a Device. Cognitive Science 8, 255–273 (1984)
4. Psotka, J., Massey, L.D., Mutter, S.A.: What mental model should be taught: Choosing Instructional Content for Complex Engineered systems. In: David, E.K. (ed.) Intelligent Tutoring Systems: Lessons Learned, pp. 85–112. Lawrence Erlbaum, Mahwah (1988)

A Solution to Revisitation Using Organic Bookmark Management

Siu-Tsen Shen[1], Stephen D. Prior[2], and Kuen-Meau Chen[3]

[1] Department of Multimedia Design, National Formosa University, 64 Wen-Hua Rd, Hu-Wei 63208, Taiwan
[2] Department of Product Design and Engineering, Middlesex University, London N14 4YZ, United Kingdom
[3] Department of Industrial Design, National United University, 1 Lien Da, Kung-Ching Li, Maioli 36003, Taiwan
stshen@nfu.edu.tw

Abstract. This research paper presents the design and user evaluation of an add-in software program referred to as Organic Bookmark Management (OBM). This system will complement the Bookmark and History functions by enabling users to navigate more efficiently using organic visual graphical cues. The findings from formative user studies conducted by this research have defined web usage and analysis of web browsing in terms of navigation patterns. Evaluation of the OBM alternative to the normal "hub and spoke" navigation structure of traditional Bookmarks and History functions will be conducted. The main difference between this schema and conventional designs is that it maintains a complete and consistent visual display of previously bookmarked and visited pages based on an organic metaphor. Implementation decisions and present results of usability studies in which we deploy the prototype are discussed. The results show that OBM brings qualitative improvement to the browsing experience of users.

Keywords: Web browser, revisitation, re-finding, organic bookmark management.

1 Introduction to Web Browsers

Currently the most popular browsers are Microsoft Internet Explorer (46.9% market share worldwide), Mozilla Firefox (30.8%), Google Chrome (14.9%), Apple Safari (4.8%) and Opera (2.1%) (see Fig.1). It is interesting to note that Microsoft's Internet Explorer browser had fallen below 50% of worldwide market share for first time in September 2010 [1] and as of January 2011, it has been surpassed by Firefox in the European market.

Revisitation is one of the most important user browsing behaviors, requiring an individual's skill to retrace their steps in order to find a piece of information. The massive data flows available and the ubiquity of the web makes this task all the more difficult. With one third of the world's population now able to access the Internet, there are over two billion Internet users with a similar shared problem (see Fig. 2).

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 46–52, 2011.

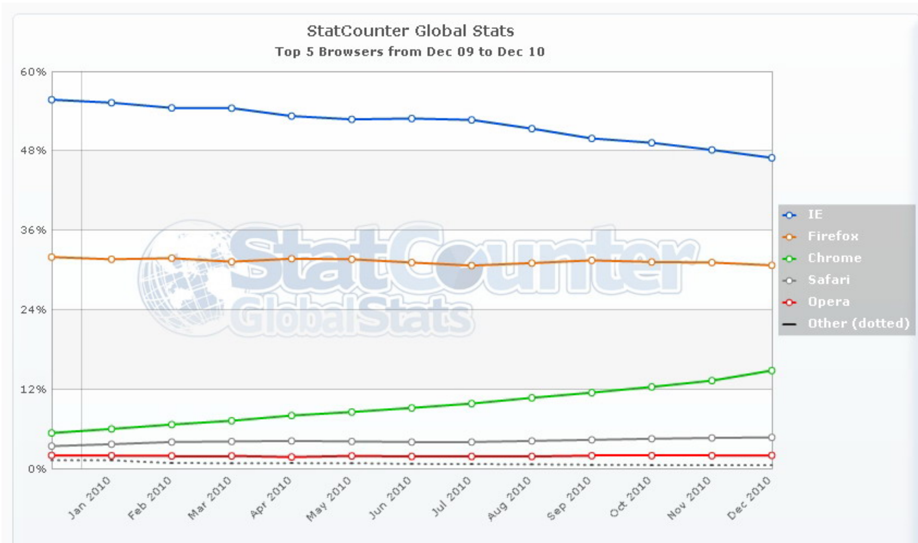


Fig. 1. Top 5 browsers by market share [1]

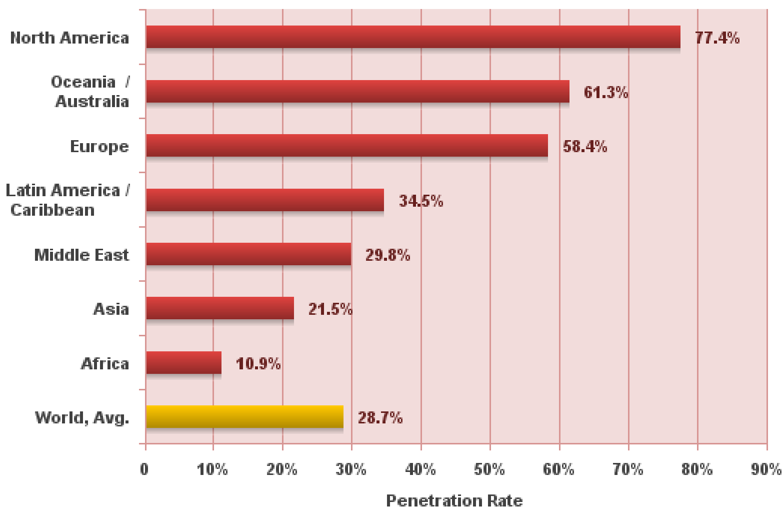


Fig. 2. World Internet Penetration Rates by Geographical Region [2]

The growth and use of the internet over the last two decades has transformed the way the world works, learns and plays. However, the sheer volume of available data has resulted in information overload. The ability to store and retrieve information for use at a later date is therefore a crucial tool which browser software developers have struggled with since their introduction in the early 1990's.

2 Literature Review of Bookmarks and History Management

Research has shown that 81% of web pages have been revisited [3]. Bookmarks and History are two key functions to support users in revisitation. Bookmarking demands individual effort to add-in and organize web data, while History offers to re-find the previously visited web sites based on the users records. Nevertheless, the maintenance of Bookmarks has been one of the top three usability problems on the web [4]. This could be explained by the fact that the system of Bookmarks management depends largely on tagging mechanisms and efficient directory order. As long as the tags or directory are not well organized, it is very difficult to retrieve information. The use of visual linear historical lists in the History function has been proven not helpful. A sense of disappointment often occurs when a desired web page could not be traced or when the user is unable to re-find info is older than a specific period.

Several research groups have looked at improving the Bookmarks and History functions, such as Data Mountain's visual screenshots of visited sites and PadPrints' graphical history map of visited pages [5, 6]. Kaasten and Greenberg (2001) even suggested combining Back, History and Bookmarks in web browsers [7]. However, all these attempts have shown their limitations to support revisitation. Cockburn and Mckenzie (2001) pointed out that there is still the lack of empirical evaluation in revisitation, especially when there are more and more users who tend to build larger bookmark collections, and this leads to the current interface schemes that produce extremely long textual menus, and therefore force users to re-organize their bookmark structure by their own efforts [3].

Currently, web browsers still rely on the individual's capability and willingness to organize their bookmarks and history. Therefore, the question remains, how to assist users with visual management of data in a more logical and easily-manipulated

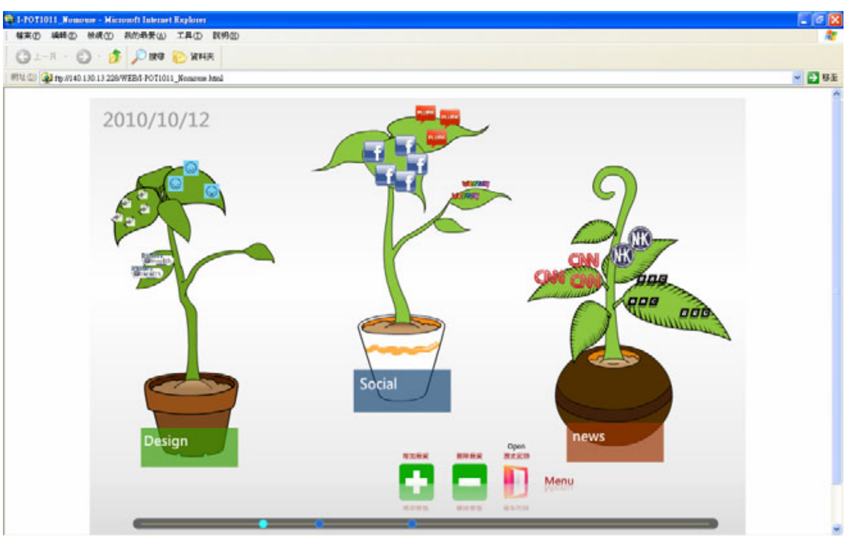


Fig. 3. Screenshot of the Organic Bookmark Management Schema

method? It is said that users can remember better with recalled attributes such as the file location, file type, keywords, associated events e.g. text messages and emails, and visual elements [8]. Furthermore, some studies suggest that the use of thumbnail previews is helpful in searching [7, 9]. Dziadosz and Chandrasekar (2002) showed that the combination mode of text summaries and thumbnail previews is better than text-only summaries and thumbnail previews only [10].

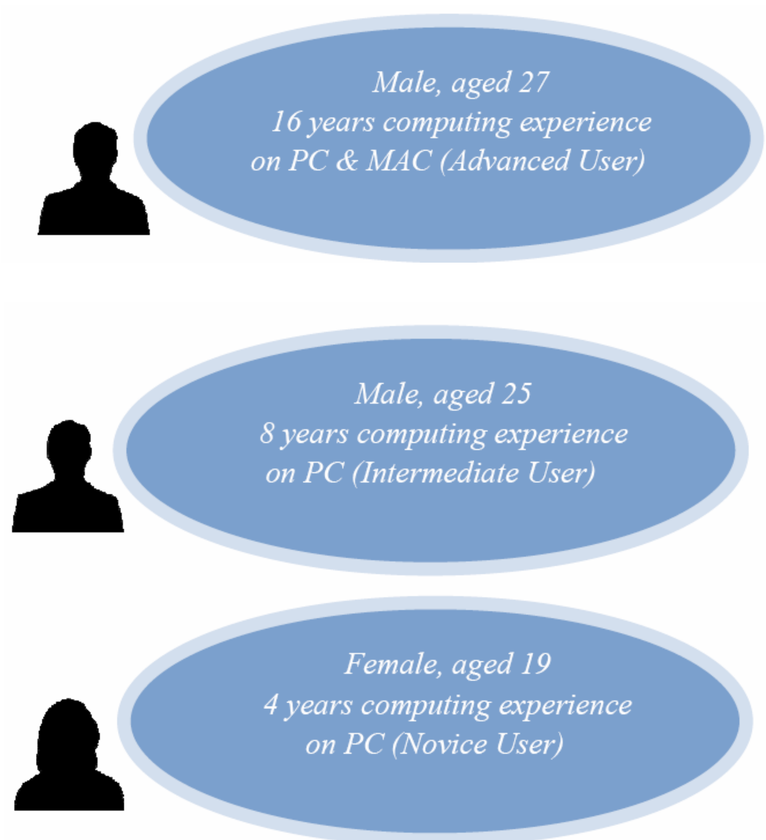
Hopefully, these valuable results will be incorporated within a new approach i.e. Organic Bookmark Management (OBM) that could simulate the growth of vegetation to facilitate revisitation. The OBM visualization aims to make information retrieval more efficient and effective by easily organizing and managing online data. By integrating the application into the web browser, OBM can build on existing functionality to trace the visited pages and reconstruct an organic look and feel.

3 Research Method

Formative studies were conducted between September and October 2010 at the College of Art and Humanities, National Formosa University (NFU). The purpose was to obtain a qualitative understanding of how an individual performed during browsing activity, and what level of effort was required to organize and manage their History and Favorites (Bookmark). The study consisted of 40 participants (20 male, 20 female) in total, who were given 800 NT$ (approx. 25 US$) individually. The process started with gathering user's background information; then was followed by further general questions relating to the use of the History and Favorites (Bookmark) functions. The basic background information included: (a) gender: *males vs. females*; (b) educational background: *high school vs. college vs. university vs. higher qualification*; (c) level of computer experience: *0-2 yrs vs. 2-4 yrs vs. 4-6 yrs vs. 6-8 yrs vs. 8-10 yrs vs. >10 yrs*; (d) daily use of the Internet: *0-2 hrs vs. 2-4 hrs vs. 4-6 hrs vs. 6-8 hrs vs. 8-10 hrs vs. >10 hrs*; (e) the user's regular computer platform: *PC vs. Mac*; and (f) type of web browser of daily use: *IE vs. Firefox vs. Safari vs. Chrome vs. Others*. These six variables were tested independently to evaluate its overall usability. The initial results have already been published in detail; we therefore prefer to concentrate here on a discussion of the second phase of the usability study [11]. Within the second part of the study, a small sample of the participants was selected to do further recall and recognition testing on the Organic Bookmark Management (OBM) prototype based on their personal History and Favorites (Bookmark) data submitted during the initial screening.

4 Discussion of the Results

To determine if the Organic Bookmark Management (OBM) prototype could visually support users' revisitation or not, we therefore selected three participants from the group which included one novice, one intermediate, and one advanced user to evaluate its usability. The reason for choosing such a small number from amongst the 40 participants is because we would like to gain the general scope of the user pattern across different levels of computer skills in browsing, especially for the first stage in developing the OBM prototype. This group consisted of two males (advanced and intermediate), and one female (novice); see below for more details of their experience and backgrounds.



In terms of the number of bookmarks, the Advanced user had more than 200, the Intermediate user had over 100 and the Novice user had less than 10.

A mock-up of the OBM prototype (developed in Flash) was given to them to try and test in their own time for a week at home. After this period, they returned to the University where they were asked a series of questions in an informal discussion. There were six main questions relating to the OBM prototype: (Q.1) the ease of understanding the OBM metaphor; (Q.2) the easiness with which to manipulate the OBM Favorites (Bookmark) function; (Q.3) the easiness with which to view the History function; (Q.4) the usability of organizing the OBM system; (Q.5) the helpfulness of the OBM timeline; (Q.6) further suggestions to improve the OBM system.

Answers to the first five questions above were ranked using a five-point Likert scale (5=Totally Agree, 4=Agree, 3=Neutral, 2=Disagree, and 1=Totally Disagree) and can be seen in Fig. 4 below.

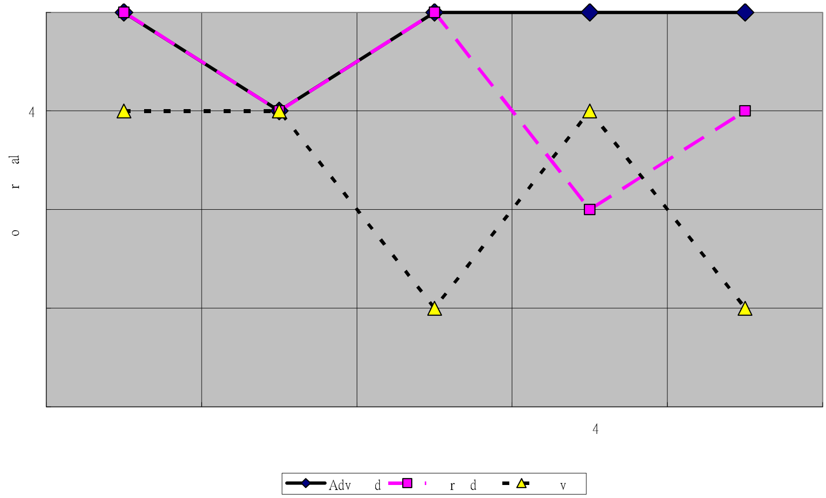


Fig. 4. Usability Evaluation Part II for OBM Prototype

Overall, all three users were generally satisfied with the look and feel of the OBM prototype. However, as can been seen above, the novice user found aspects of the History function and the timeline less than satisfactory. Clearly, the number of years of computer experience has an effect on the level of positive feedback WRT the OBM prototype. Regarding Q.6, the users further suggested how to improve the OBM system by adding more functions such as RSS Reader and Feed for further usability testing.

5 Conclusions

Analysis of our data shows that most participants were regarded as expert users with more than 10 years computer experience. Over half the participants surfed the Internet for more than 8 hours a day. The vast majority of participants were PC users. It is interesting to note that our participants (both male and female) were broadly in line with the worldwide statistics for the use of web browsers. Even though all the participants were aware of the History function, there were still 40% of the males and 65% of the females who did not regularly use it. This was overwhelmingly due to it being perceived as unhelpful, confusing or difficult to use. Almost all the participants were aware that the History function could display all the websites that were visited in the last few weeks. Nevertheless, 25% of the males and 45% of the females did not know that they could change the default History duration from 21 days to whatever they liked. The majority of participants knew the level of functionality within the History function; however, four times as many women (20%) as men (5%) did not. Moreover, 25% of the males and 50% of the females did not know that you could search the History function using a keyword.

There appears to be a gap in knowledge between the males and the females as to some of these basic functions. Is it that they never learnt about them, or is it that they simply don't find them easy to use and therefore they do not use them at all?

All of the participants were aware of the Favorites (Bookmark) function and almost all of the participants used this function. Nevertheless, almost all the participants found it difficult to manage their Favorites (Bookmarks). When asked about the problem of revisiting a web page that they knew they had been to in the recent past, 60% of the males and 65% of the females stated that they often experienced problems with this. The level of acceptability for users to adopt the OBM system over the traditional offerings from Internet Explorer and Firefox remains challenging. However, an add-in software extension to IE and Firefox would seem practical and beneficial.

The results of this formative study provide a solid foundation for future development of the OBM system. We believe that even the most successful browser, i.e. Internet Explorer can be improved by redesigning their Favorites and History functionality.

References

1. StatCounter. Microsoft's Internet Explorer browser falls below 50% of worldwide market for first time (2011), `http://gs.statcounter.com/press/microsoft-internet-explorer-browser-falls-below-50-perc-of-worldwide-market-for-first-time` [cited January 22, 2011]
2. Group, M.M. Internet Usage Statistics. World Internet Users and Population Stats (2011), `http://www.internetworldstats.com/stats.htm` [cited January 22, 2011]
3. Cockburn, A., McKenzie, B.: What Do Web Users Do? An Empirical Analysis of Web Use. International Journal of Human-Computer Studies 54(6), 903–922 (2001)
4. Pitkow, J. GVU's WWW User Survey (1996), `http://www.cc.gatech.edu/gvu/user_surveys/survey-04-1996/` (cited)
5. Robertson, G., et al.: Data Mountain: Using Spatial Memory for Document Management. In: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. ACM Press, San Francisco (1998)
6. Hightower, R., et al.: Graphical Multiscale Web Histories: A Study of Padprints. In: Proceedings of the ACM Conference on Hypertext. ACM Press, New York (1998)
7. Kaasten, S., Greenberg, S.: Integrating back, history and bookmarks in web browser. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems. ACM, Seattle (2001)
8. Blanc-Brude, T., Scapin, D.L.: What do People Recall about their Documents? Implications for Desktop Search Tools. In: Proceedings of Intelligent UserInterfaces 2007. ACM, Honolulu (2007)
9. Czerwinski, M., et al.: The Contribution of Thumbnail Image, Mouse-over Text and Spatial Location Memory to Web Page Retrieval in 3D. In: Proceedings of Interact 1999 (1999)
10. Dziadosz, S., Chandrasekar, R.: Do Thumbnail Previews Help Users Make Better Relevance Decisions about Web Search Results? In: Proceedings of SIGIR 2002. ACM, Finland (2002)
11. Shen, S.-T., Prior, S.D., Chen, K.-M.: Revisiting Revisitation in Computer Interaction: Organic Bookmark Management. In: Fifth International Conference on Design Principles and Practices. Common Ground Publisher, Sapienza University of Rome (2011) ISBN1833-1874

A Study on the Time Estimation Measurement for Web Usability Evaluation

Keiji Suzuki, Mitsuhiko Karashima, and Hiromi Nishiguchi

2-3-23 Takanawa, Minato-ku. Tokyo, Japan
mitsuk@tokai-u.jp

Abstract. In this research the effectiveness of the time estimation as the measure for the efficiency relating to the cognitive workload, which was a component of the usability, was examined through the usability tests experiment. Sixteen subjects were required to carry out two tasks according to the two scenarios with the low and high usability websites of the local governments. The result of time estimation revealed that the efficiency of the high usability website was higher than the low usability website, the same as the results of the other measures. From the results of this experiment it was suggested that the time estimation could be an effective measurement for the efficiency relating to the cognitive workload which was a component of the usability.

Keywords: usability, cognitive workload, time estimation.

1 Introduction

In ISO9241-11[1] usability is defined "Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". The effectiveness is defined as "Accuracy and completeness with which users achieve specified goals". The efficiency is also defined as "Resources expended in relation to the accuracy and completeness with which users achieve goals". The satisfaction is defined as "Freedom from discomfort, and positive attitudes towards the use of the product".

ISO 9241-11 also describes the usability measurements. It describes that relevant resources to efficiency can include mental or physical effort, time, materials or financial cost, and that, for example, human efficiency could be measured as effectiveness divided by human effort, temporal efficiency as effectiveness divided by time, or economic efficiency as effectiveness divided by cost. In the annex B it is also described that the cognitive resources expended in the conduct of tasks should be measured as one of measures of human efficiency.

However the measures of the efficiency relating to the cognitive workload were a few except the task time, the NEM (Novice Expert ratio Method) by Urokohara. (1999)[2], and several subjective Mental Work Load (MWL) measures, for example, Subjective Mental Effort Questionnaire (SMEQ) (Zijlstra, 1993[3]), and NASA-TLX((NASA-Ames Research Center, 1986[4])) (Nigel, 1995[5]), and so on.

The literatures suggested that the difference between the task time and the time estimated by the time estimation measures could reflected the Central Processing

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 53–59, 2011.

Load (CPL) which meant the degree of the working memory resources consumption (Brown, 1997[6], Shinohara, 2002[7]). The higher the degree of the working memory consumption was, the more slowly the time passed in one's brain, and the estimated time became shorter. It suggests that the time estimation measures should be able to be the measure of the efficiency relating to the cognitive workload. Mats and Henning (2007)[8] tried to access the effectiveness of the time estimation measurement as an indicator of MWL and suggested that the time estimation measurement could estimate the difference of the MWL between the tasks which included finding specific pieces of information on websites, solving computer based puzzles, and solving a simple sudoku game. These researches suggest that the time estimation measurement might become a non-subjective measure of the efficiency relating to the cognitive workload which is a component of the usability.

This research focused on the time estimation measurement and the effectiveness of the time estimation measurement as the measure of the efficiency relating to the cognitive workload, which was a component of the usability, was examined through the usability tests' experiment.

2 Methods

2.1 Subjects

Sixteen male students participated in this experiment as the novices who had no experience of two websites which were prepared for the experiment and were divided into two groups. Each group was consisted of eight subjects. Four other graduated students participated in this experiment as the experts who knew the place of information on the websites well but they had no experience of operating the websites.

2.2 Experimental Tasks

Sixteen novice subjects were required to carry out two tasks according to the two scenarios with one website of the local government as the usability test of the website. The subjects of one group used the website which won the high rank at the HCD-net's local government website usability award (Web1) and the subjects of the other group used the website which was unranked (Web2) [9].

The subjects were required to find two kinds of the information on the website for moving house which two scenarios indicated. The subjects were required to find information on the website to have the government collect large size waste (Task1). The subjects were also required to find how to register for the national health insurance (Task2).

Four expert subjects were required to carry out the two tasks with the two websites.

Sixteen novice subjects were also required to reproduce the time spent in the task by using the stopwatch after carrying it out. As the method of time estimation the prospective paradigm was adapted. In the prospective paradigm the subjects were given notice to reproduce the time spent before it was experienced.

2.3 Apparatus

The subjects performed the tasks with the FireFox browser on Windows XP. A PC was equipped with a 17 inch CRT monitor and connected to the Internet. No time related information was visible on the monitor. The subjects were asked to remove any time keeping device. Every task by every subject was recorded by the digital video camera and the time spent in finding the required information from the website (the task time) was measured later from these recordings.

2.4 Experimental Procedure

As the control of the time estimation, sixteen novice subjects experienced 20, 30, and 40 seconds on the seats with eyes open and they were required to reproduce the time and to answer the subjective MWL through the NASA-TLX for time estimation. They were required to carry out one of two tasks according to the scenario with the given website, to reproduce the time spent in the task after carrying out it, and to answer the subjective MWL for the task with time estimation through the NASA-TLX. They were also required to carry out the other of the tasks, to reproduce the time, and to answer the NASA-TLX. The order of carrying out two tasks was counterbalanced by the subjects.

Four expert subjects were only required to carry out the two tasks with the two websites. The orders of carrying out two tasks and using two websites were counterbalanced by the subjects.

2.5 Measures

The task time was measured. The time which the subjects reproduced (the estimated time) was also measured. In the prospective paradigm, if the working memory resources consumption is more, the time passed more slowly in the subject's brain and the estimated time is shorter.

The ratio of the task time by the novice subject to the averaged task time by the expert subjects was calculated as the NEM score. The ratio of the task time to the estimated time (*R(te(task time))*) was calculated according to the equations (1) and (2) as the index of the time estimation. The higher this ratio was, the time passed more slowly in one's brain.

$$te(actual\ time) = \frac{actual\ time}{estimated\ time} \tag{1}$$

$$R(te(task\ time)) = te(task\ time) \Big/ \frac{te(20\,\mathrm{sec}) + te(30\,\mathrm{sec}) + te(40\,\mathrm{sec})}{3} \tag{2}$$

The AWWL (Adaptive Weighted WorkLoad) score of NASA-TLX which was proposed by Miyake (1993) [10] was also calculated as the index of the subjective MWL.

3 Results

The result of two-way mixed ANOVA for the task time of sixteen novice subjects in tasks and webs revealed that the task time for Web2 was significantly longer than Web1 regardless of the tasks($F(1,14)=9.58$ $p<0.01$) as shown in Fig.1..

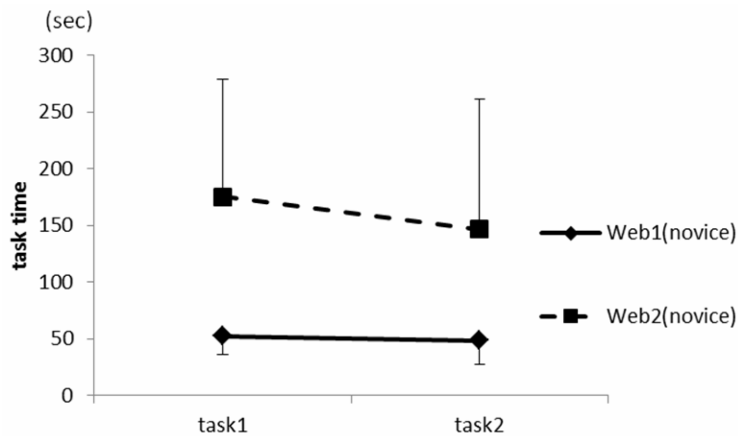


Fig. 1. Task time (novice subjects)

On the other hand, the result of two-way within-subjects ANOVA for the task time of the expert subjects in tasks and webs revealed that there was no significant difference in the task time between the webs regardless of the tasks ($F(1,6)=0.435$, $p>0.10$) as shown in Fig.2.

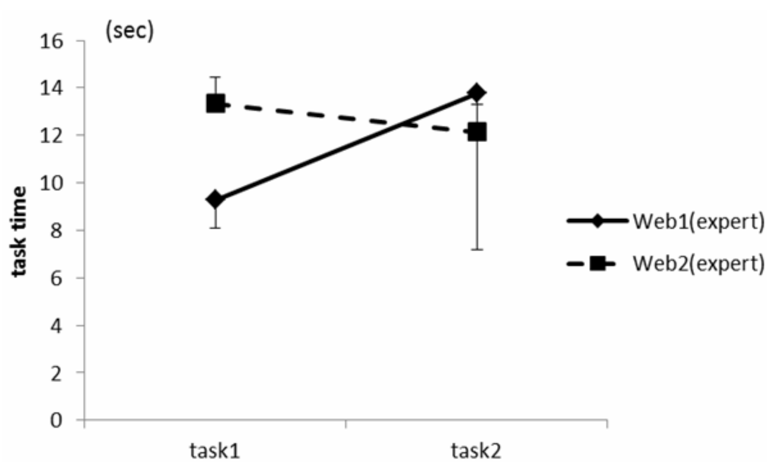


Fig. 2. Task time (expert subjects)

The result of the two-way mixed ANOVA for the NEM score in tasks and webs revealed that the ratio for Web2 was higher than Web1 regardless of the tasks (F(1,14)=8.04, p<0.05) as shown in Fig.3.

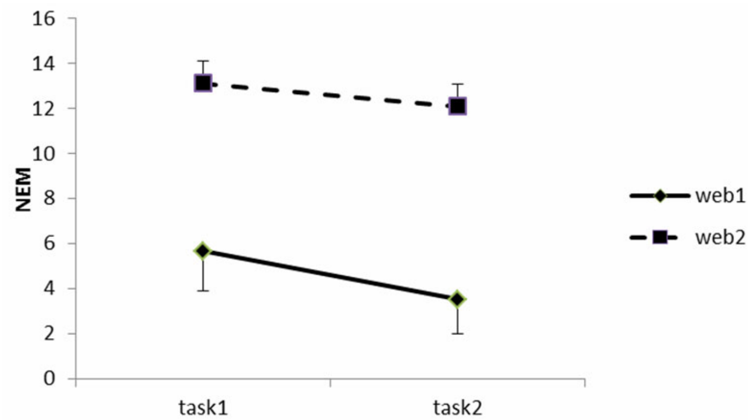


Fig. 3. NEM

The result of the two-way mixed ANOVA for the AWWL score in tasks and webs revealed that the score for Web2 was slightly higher than Web1 regardless of the tasks (F(1,14)=3.65, P<0.10) as shown in Fig.4.

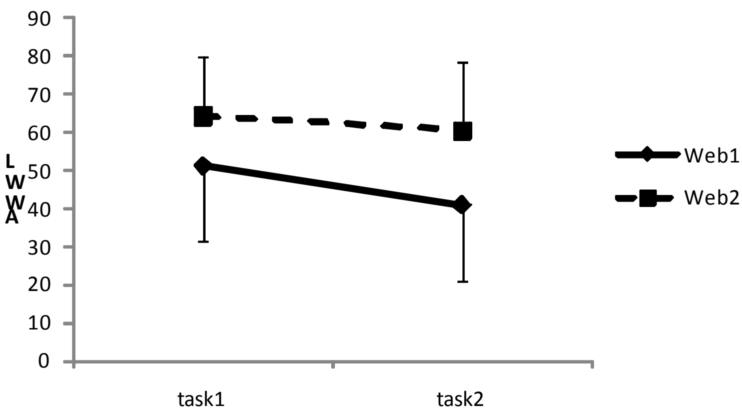


Fig. 4. AWWL

The result of two-way mixed ANOVA for the ratio of the task time to the estimated time(*R(te(task time))*) revealed that the ratio for Web2 was significantly higher than Web1 regardless of the tasks (F(1,14)=4.92, p<0.05) as shown in Fig.5.

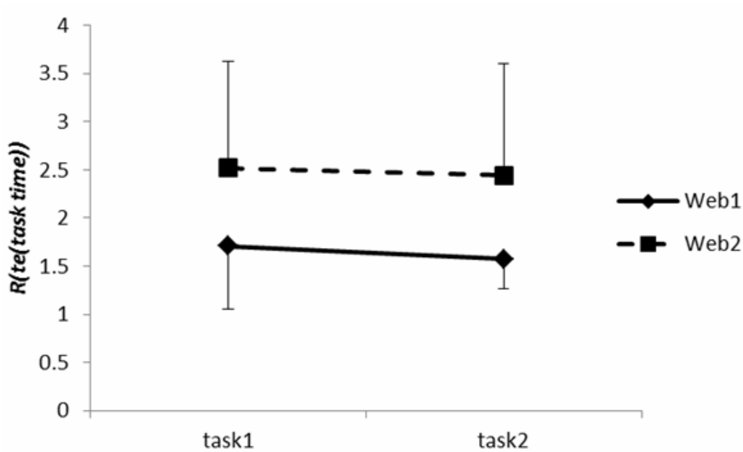


Fig. 5. Time Estimation

4 Discussions

The result of the task time of the novices revealed that the spent time in finding information on Web2 was longer than on Web1 regardless of the tasks. This result meant that the difficulty of finding information on Web2 or operating Web2 were higher than Web1.

On the other hand the result of the task time of the experts revealed the spent time of finding information on Web2 was the same as on Web1 regardless of the tasks. The result of the task time of the experts suggested that the difficulty of operating Web2 was the same as Web1 because the difficulty of finding information on Web2 was the same as Web1. As the experts had no experience of operating the websites the same as the novices, the genuine operating time might not influence the difference of the task time of the novices between the websites.

From these results of the task times it was suggested that the difficulty of finding information on Web2 was higher than on Web1 and that the efficiency relating to the cognitive workload for Web1, which was a component of the usability, was higher than Web2.

The result of the NEM revealed that the score of Web2 was higher than Web1. It suggested that the difficulty of finding information on We2 was higher than on Web1 and that the efficiency relating to the cognitive workload for Web1 was higher than Web2 as the difficulty of the operating Web2 was the same as Web1.

The result of the NASA-TLX revealed that the AWWL of Web2 was slightly higher than Web1. It suggested that the subjective mental workload of Web2 was slightly higher than Web1.

These results of the indicators confirmed that the efficiency relating to the cognitive workload for Web1, which won the high rank at the HCD-net's local government website usability award, was higher than Web2 which was unranked.

The results of the time estimation revealed that the time passed in the subject's brain more slowly when he was finding information on Web2 than on Web1. This result of time estimation meant that the degree of the working memory resources consumption for Web1 was less than Web2, the same as the results of the other measures that revealed the efficiency relating to the cognitive workload for Web1 was higher than Web2.

From these results of this experiment it was suggested that the time estimation could be the effective measure for the efficiency relating to the cognitive workload which was a component of the usability.

5 Conclusion

In this research the effectiveness of the time estimation measurement as the measure of the efficiency relating to the cognitive workload, which was a component of the usability, was examined through the usability tests experiment. From the results of this experiment it was suggested that the time estimation could be the effective measure for the efficiency relating to the cognitive workload, which was a component of the usability.

Acknowledgements. This research was supported by Grant-in-Aid for Scientific Research(c), 2009-2011, 21510162 from Japan Society for the Promotion of Science.

References

1. ISO9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) -Part 11: Guidance on usability (1998)
2. Urukohara, H.: Method to build good usability: Task analysis and user interface design using operation flowcharts. In: Human-Computer Interaction, vol. 1, pp. 928–932. Lawrence Erlbaum, London (1999)
3. Zijlstra, F.R.H.: Efficiency in Work Behaviour: a Design Approach for Modern Tools. Delft University Press, Delft (1993)
4. NASA-Ames Research Center, Human Performance Group: Collecting NASA Workload Ratings: A Paper-and-Pencil Package. NASA-Ames Research Center, Moffet Field (1986)
5. Nigel, B.: Measuring usability as quality of use. Software Quality Journal 4, 115–150 (1995)
6. Brown, S.W.: Attentional Resources in Timing: Interference Effects in Concurrent Temporal and Nontemporal Working Memory Tasks. Percept. Psychophys. 59, 1118–1140 (1997)
7. Shinohara, K.: The role of working memory cognitive process of time estimation. Psychological Theory 45(2), 195–209 (2002)
8. Mats, L., Henning, S.: Time Estimation as a Measure of Mental work load. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS (LNAI), vol. 4562, pp. 359–365. Springer, Heidelberg (2007)
9. The report of the HCD-net's local government website usability award (in Japanese), `http://www.hcdnet.org/pdf/WEBevaluation2007.pdf`
10. Miyake, S., Kumashiro, M.: Subjective mental workload assessment technique: An introduction to NASATLX and SWAT and a proposal of simple scoring methods. The Japanese Journal of Ergonomics 29(6), 399–408 (1993) (in Japanese)

Study of User Interface for Browsing Web Contents That Considers the Cognitive Features of Older Users

Masahiro Watanabe, Shunichi Yonemura, Ryo Hashimoto, and Yoko Asano

Cyber Solutions Laboratories, Nippon Telegraph and Telephone Corporation,
1-1 Hikarinooka, Yokosuka, Kanagawa, 239-0847, Japan
{watanabe.masahiro,yonemura.syunichi,hashimoto.ryo,
asano.yoko}@lab.ntt.co.jp

Abstract. Web accessibility for old users has become a serious issue, especially in Japan. The problems involve cognitive as well as physical characteristics. Cognitive problems are well-handled by the metaphor approach, especially for older users when Web browsing. In order to investigate the impact of his approach, we conducted experiments with 11 old subjects and 10 young subjects. They were asked to search for a target in a Web site via the book metaphor interface and with a common Web browser interface. Although there were no differences in the task success rates or the task completion time between the two interfaces, there was a difference in the browsing time per Web page. The results of a questionnaire show that many old users prefer the metaphor interface. With the book interface, they selected the strategy of look and click because it lessens the demands placed on working memory.

Keywords: older user, Web, accessibility, usability, book metaphor.

1 Introduction

The rapid penetration of the Internet has made Web sites one of the most important ways of passing and acquiring information today. The relative newness of Web site design principles has created serious usability issues, particularly for people with disabilities and older users; this is especially true in Japan because it has already become an aged society. It is important to focus not only on the physical issues of the elderly but also on their cognitive issues. A lot of research is targeting Web accessibility, the ability to access information [1 - 3] and there are some guidelines for Web design [4, 5]. These studies and guidelines mainly focus on physical issues, and few Web design methods that are appropriate for the cognitive issues of older users have been published.

Many laboratories have executed psychological studies on the cognitive issues of the elderly under strictly controlled conditions. Study issues include attention and working memory, and results have shown the negative effects of aging. Although field work with existing Web sites is important, these studies did not address the usability of Web content for older users.

Some Web designs cause various troubles for old users in browsing Web sites, for example they have trouble in scrolling a Web page. These troubles are caused by the

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 60–67, 2011.

cognitive weaknesses of users, and have no known direct cause. We focused on the dialog principles of ISO 9241-110 [6], and found that self-descriptiveness is a critical dialogue principle for older users [7]. In order to satisfy the principle and to reduce the impact of the cognitive issues of older users, we think that the metaphor approach is very attractive. This is because the metaphor represents something they are familiar with in daily life and so places no unusual demands on their working memory. We focused on the book metaphor as a Web site interface and investigated its usability for older users.

2 Experiments

Eleven older users (65 - 74 years old; 6 females) and 10 young users (20 - 33 years old; 5 females), who had experience of PC operation for 3 months - 15 years and use the PC less than 4 times a week, participated in the experiments. They were requested to access a Web site managed by our research group and accomplish search tasks using two interfaces: a common Web browser (Fig. 1: browser interface), and the book metaphor interface (Fig. 2: metaphor interface). In order to investigate the impact of the metaphor, we made a Web site that looked like book, see Fig. 2, mainly by changing the style sheets of the Web page. Both interfaces displayed the same information.

The subjects were instructed to accomplish 4 search tasks as shown below and not to use keyword search. We made 2 simple tasks (1-1 and 1-2) and 2 complicated tasks (2-1 and 2-2). The site consists of many pages that are categorized. In the complicated task, the subjects had to locate the keyword in one category while holding the other keyword in memory, and then to locate the second keyword in another category. This tasks places more loads on the subject’s working memory. The experiment with the complicated task followed that with the simple task. In both tasks, the assigned task order, 1-1 or 1-2 (2-1 or 2-2), was randomized between subjects.

- Task 1-1: Find out via the Internet the conference attended by 102 people in the event “NTT Group Collection 2007.”
- Task 1-2: Find out via the Internet the favorite football team of Mr. A (displayed on the page “Key person for universal design”). We used the real name of Mr. A in the experiment.
- Task 2-1: Find out via the Internet the telephone number of the NTT-West information equipment division that made the telephone terminal, ”IP Telephone UD.”
- Task 2-2: Find out via the Internet the name of 4 NTT DoCoMo staff members who created a mobile phone with 2 displays as part of a universal design activity.

We observed each subject while performing the tasks and interviewed each. The subjects were requested to select their preferred interface. The Web pages were generated within Internet Explorer 6 running on Windows XP, and presented on a liquid crystal display (21.3 inches, UXGA). The window of the browser was fixed at full screen mode.

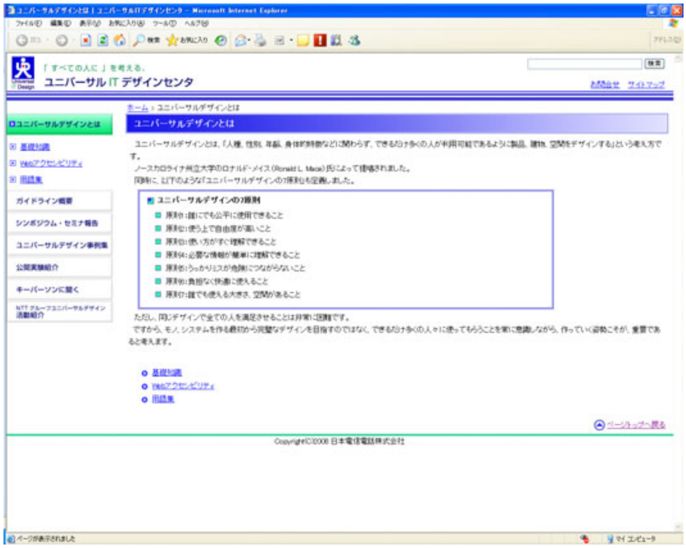


Fig. 1. Web page presentation via existing Web browser (browser interface)

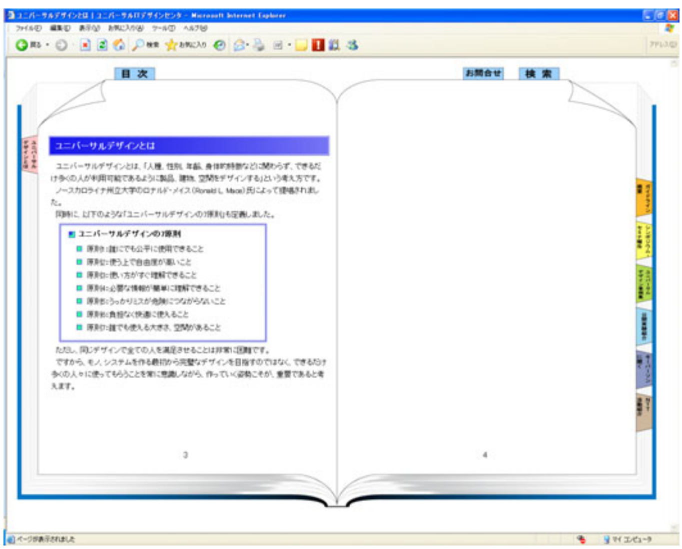


Fig. 2. Web page presentation via book metaphor (metaphor interface)

In the metaphor interface, all Web pages in the site are linked each other in one line like a real book. As shown in Fig. 2, the right and left upper corners of the book are everted and links are provided by using image maps. The image map at right upper corner links to the next page and that at left upper corner links to the previous page. Animated page turning was not used.

To provide the same navigation methods in both interfaces, we designed the 7 tabs in the metaphor interface (Fig. 2) to match the left side navigation menus in the browser interface (Fig. 1). The same information is provided in both interfaces and the sentences provided the same links. The subjects were instructed to accomplish the tasks as quickly as possible and we timed the task completion time. If the subject found the right answer within 15 minutes the trial was deemed a success, otherwise, a failure. Even if the page holding the answer was displayed, the subject had to provide the right answer. Trials were terminated if

1. Subject gave an answer.
2. Subject gave up.
3. Fifteen minutes passed.

The subjects were asked to say “I found the target.” or “I give up” to the experimenter and we stopped timing at these declarations. This means that the task completion time is the whole time of Web operation regardless of success or failure. Right before the first task in the metaphor interface, the subjects were asked to practice by freely operating the book interface for 2 minutes. After completion of all tasks, we determined the preferred interface from interviews and six questions.

3 Results

3.1 Task Success Rate

Fig. 3 shows the success rates for each subject, each task, and each interface. Task 1 (task 1-1 or 1-2) was comparatively easy, while task2 (task 2-1 or 2-2) was comparatively difficult. In all conditions, older subjects had lower success rates than the young subjects. There were no significant differences with regard to success rates between browser and metaphor interfaces.

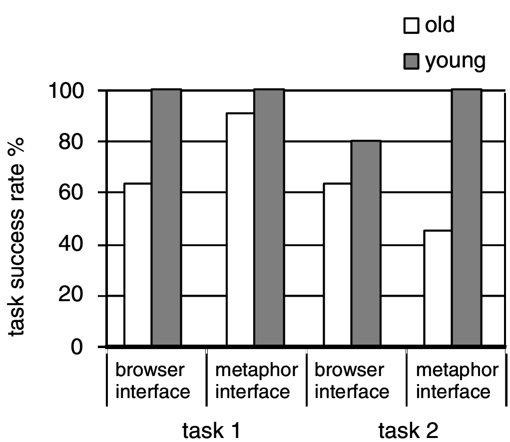


Fig. 3. Task success rate

3.2 Task Completion Time Periods

Fig. 4 shows the mean task completion times. Vertical bars show the standard deviations. The times are for the subjects who performed the task successfully. The data of subjects who failed the task was not used.

The old and young subjects demonstrated the same completion times for task 1 regardless of the interface. In task 2, although there was no difference in the time taken by old and young subjects in the metaphor interface, the old subjects took longer than the young subjects in the browser interface (Kruskal-Wallis test, $p < 0.01$). ANOVA showed that the young subjects took longer for task 1 than for task 2 of ($F = 5.34$, $p < 0.05$) in both interfaces, although the old subjects showed no such difference according to the Kruskal-Wallis test.

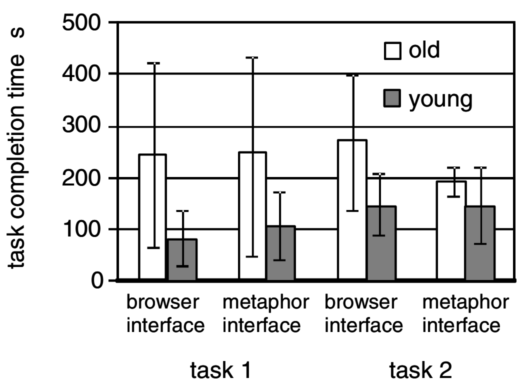


Fig. 4. Task completion time

3.3 Browsing Time per Web Page

Fig. 5 shows the browsing time per Web page. Vertical bars show the standard deviations. The browsing time per page indicates the ease with which a page can be processed. In order to investigate the subjects' search strategies, the data is taken from both successful and failed trials.

In task 1, the per page browsing time of old subjects in the metaphor interface was shorter than that in the browser interface (Kruskal-Wallis test, $p < 0.05$), although the young subjects showed no such difference. In task 2, the old subjects showed no difference in per page browsing time, while the young subjects showed shorter per browsing times in the metaphor interface than in the browser interface (Kruskal-Wallis test, $p < 0.01$).

In task 1, there was no difference between the per page browsing times of old and young subjects, but the old subjects showed longer times than the young subjects in task 2 (Kruskal-Wallis test, $p < 0.05$).

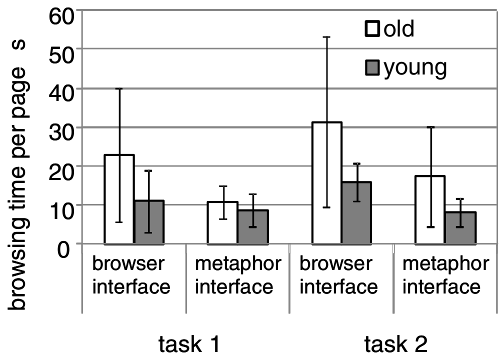


Fig. 5. The browsing time period per page

3.4 Replies to Questions and Interviews

Fig. 6 shows the subject's replies to 6 questions. The metaphor interface was preferred by old subjects in 77 % of the questions and by the young in 62 % of the questions. Over 90 % of the older users and over 60 % of the young users preferred the metaphor interface in at least one question. In interviews, some older users noted that they preferred the metaphor interface because "It is easy to learn how to use it. It is used as a book." Another said, "It looks like it contains a small amount of information." Another reply was "It does not use Web page scrolling." These answers imply that the metaphor interface does not overload the cognitive abilities and well-suits older users. We gathered from the subjects their reasons for making the choices shown in Fig.6. The major opinions are shown in Table 1.

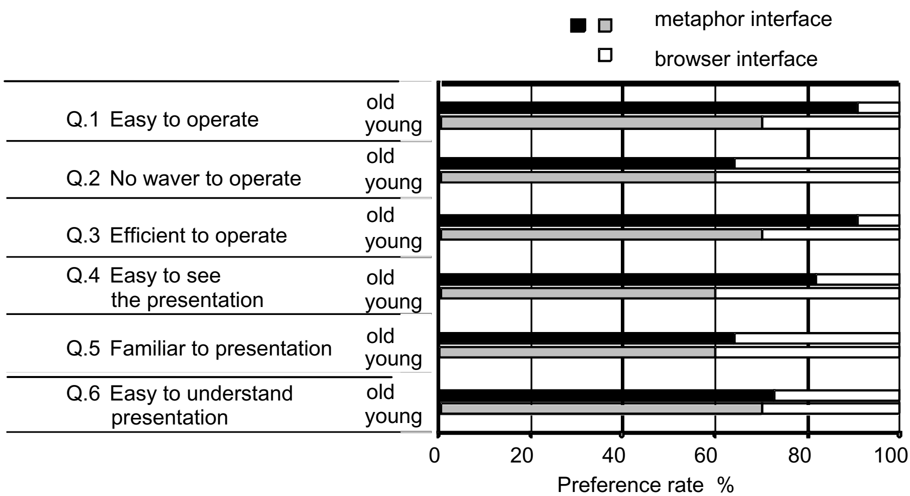


Fig. 6. Preference rates

Table 1. Major reason for preferences

Selected interface	subjects	Major reasons why they chose the interface
Select metaphor interface	old	● It is similar to a book. Because I understand how to use it, I do not have to click several places. ● Page turning. If I cannot find the way to the target, all I have to do is to turn over the page. I can search sequentially. To turn over is easier than scrolling a page. ● Each page contains a well defined amount of information.
	young	● I can use it like a book. ● It has a list of content, book marks, and tabs. ● It contains very simple information
Select browser interface	old	● I often use it. I am not familiar with the book metaphor. ● The electric book is not suitable for searches. I have an image of reading a book. ● It is easy to search using keywords.
	young	● I often use it. ● It is easy to understand because there is much information on each page. ● There is a clue in a page to find the target.

4 Discussion

There was no difference in the success rates between the browser and metaphor interfaces. The success rates in the metaphor interface were high even though the book metaphor was novel to all subjects. With the metaphor interface, the old subjects basically matched the performance of the young. A key indicator of the characteristics of the metaphor interface is the average time spent per page. The metaphor interface encouraged the strategy of rapid page jumps, the subject did not give up the task and continued the search.

The replies to the six questions clearly indicated that the elderly preferred the metaphor interface. A previous study showed that object familiarity reduces the load on the working memory of old people [8]. This is confirmed by the subjects who replied that the metaphor interface was "easy to understand how to use". The metaphor interface is familiar to even novice users. All subjects had experience in reading books and could put their experience to use in browsing Web pages in the metaphor interface.

Most of the older users preferred the book metaphor. The result shows that a metaphor has a possibility to ease cognitive load of older users although there were no differences in the success rates and the performing task time period. The metaphor enhances operating Web pages, saves working memory's load and provides some sense of ease.

The metaphor interface was more popular with the old subjects than the young. The former expressed the opinion that it was easier to mentally process a block of information that had clearly defined limits, the page had borders. That is, they could see all the information on the page without additional operations like page scrolling.

While the elderly are trying to comprehend a block of text, it is counterproductive to force them to perform additional operations.

As described above, our comparison of browser and metaphor interfaces in the field of human-machine interface and information structure has shown the possibility of the latter reducing the load placed on the working memory of elderly users. It indicates the metaphor interface is one of the most suitable interfaces for older users. In order to apply this finding to practical Web design, further study is needed.

5 Conclusions

In this study, we focused on accessing the Web via the metaphor interface as a way of reducing the cognitive loads placed on the working memory of older users. A comparison of metaphor and browser interfaces showed that there was no difference in task success rate or in task completion time. Even though all subjects were experiencing the metaphor interface for the first time, their performance was not significantly degraded. It indicates that the metaphor interface is suitable for novice users. The browsing time period per page in the metaphor interface is shorter than that in the browser interface. It indicates that the interfaces yield different search strategies. Most of the older subjects preferred to book metaphor which suggests that the metaphor interface placed less loads on working memory. Future works include investigating which feature of the metaphor impacts the cognitive issue of older users and Web design methods including system design.

References

1. Paciello, M.G.: Web Accessibility for People with Disabilities. CMP Books, Lawrence (2000)
2. Thatcher, J., Burks, M.R., Heilemann, C., Henry, S.L., Kirkpatrick, A., Lauke, P.H., Lawson, B., Regan, B., Rutter, R., Urban, M., Waddell, C.D.: Web Accessibility: Web Standards And Regulatory Compliance. Apress, Berkeley (2006)
3. Nielsen, J.: Designing Web Usability: The Practice of Simplicity. New Riders, Berkeley (2000)
4. World Wide Web Consortium, Web Content Accessibility Guidelines 2.0 (2008), `http://www.wcag2.0.com/`
5. Japanese Industrial Standards Committee, Japanese Industrial Standards X 8341-3:2010, Guidelines for older persons and persons with disabilities – Information and communications equipments, software and services – Part 3: Web content, Japanese Standards Association, Tokyo (2010)
6. ISO 9241-110; Ergonomics of human-system interaction - Part 110: Dialogue principles (2006)
7. Watanabe, M., Yonemura, S., Asano, Y.: Investigation of Web Usability based on the Dialogue Principles. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 825–832. Springer, Heidelberg (2009)
8. Whiting, W.L., Smith, A.D.: Differential age-related processing limitations in recall and recognition tasks. Psychology and Aging 12, 214–216 (1997)

Exploring Cultural Variation in Eye Movements on a Web Page between Americans and Koreans

Changwoo Yang

Valdosta State University, MLIS Program,
1500 N. Patterson St. GA 31698
cyang@valdosta.edu

Abstract. This study explored differences in eye movement on a Web page between members of two different cultures to provide insight and guidelines for implementation of global Web site development. More specifically, the research examines whether differences of eye movement exist between the two cultures (American vs. Korean) when viewing a Web page, and if so, whether their eye movements are affected according to the level of Web page complexity. This study employed eye tracking methods and several eye movement metrics were measured.

Keywords: eye movement, cultural differences, human-computer interaction, web design cultural cognition.

1 Introduction

Web sites are developed in many different cultures (countries) all over the world and are open to not only domestic users but also a wide range of international users. These users react to Web sites in a different ways while navigating Web sites. Some users may suddenly feel uncomfortable and overwhelmed so they leave immediately, or give more attention to a certain design element at first glance.

Much previous research argues that Web site design and interface elements appear to be influenced to some degree by the culture in which Web sites originate [1,2,3,4,5] and there is some evidence that users performed better in a Web site culturally orientated for them [6]. Nisbett et al. [7] also argue that Westerners, in particular North Americans, attend more to focal objects, whereas East Asians attend more to contextual information.

Although not every user has the same perception and interpretation of the information contained in Web sites, different cultural groups may have different cognitive styles related to visual and multimedia factors. These differences may affect how users respond to and use online material. People from different cultures may have different levels of attention to the various visual elements or areas of a Web page. Therefore, one of the greatest challenges for Web developers is developing cross-cultural Web sites concerned with not only language, but with the cognitive aspects inherent in cultural diversity [6].

One method for exploring visual and cognitive perception is the use of eye-tracking technology. Although all cognitive differences between cultures cannot be

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 68–76, 2011.

explained by examining eye movement, there is some possibility of explaining underlying mechanisms in cognitive processing by examining patterns of eye movement between two different cultures [8]. This study explores whether there are cultural variations in eye movements on Web pages between members of two different cultures, American and Korean, in order to suggest a cognitive basis for cross-cultural Web design.

2 Methodology

2.1 Measurement of Eye Movement

The following six eye movement measures and one performance variable were measured: (1)Total fixation time, (2) total gaze time in each AOI, (3) fixation count on each area of AOI, (4) time to first fixation and first fixation area, (5) Fixation order, and (6) fixation transition.

In this study, a fixation defined as a series of samples within a 30 pixel radius for at least 100 msec.

2.2 Apparatus and Stimuli

The Tobii 1750 eye tracking system was used. This system tracked both eyes simultaneously, and sampled the user's eye position every 20msec. Accuracy of gaze estimation was 0.5 degree.

In this study, home pages or internal Web pages activated from global Web sites that have multiple language versions for each country were used as stimuli. Three Web pages in different levels of complexity for each country were modified based on existing global Web sites. Each pair of Web pages was modified to have the same layout and content with different languages (English and Korean). Although a Web page may be modified, it cannot be exactly identical due to the characteristics of language that create differences in white space, length of a sentence, or length of a paragraph. The Web pages looks like live Web pages. However, the Web pages consisted of only image files, and there are no clickable links on the pages in order to prevent from unnecessary clicks during the tasks.

This study is not to examine how specific variables affect patterns of eye movement, but rather to study how users actually see Web pages that consist of various Web design elements. Therefore, using current real Web pages is more appropriate for the purpose of this study.

2.3 Participants

Nineteen American graduate and undergraduate students (13 males, 6 females) and 19 international Korean graduate students (11 males, 8 females) at Florida State University were participated. The average age of the Korean participants was 33 with a range from 26 to 38. The average age of the American participants was 28 with a range from 20 to 43. All The Americans except one participant who moved to the US at age 2 and the Korean participants were born in their native country and were raised there at least until they earned their high school diploma.

2.4 Procedure

Three phases of tasks were performed: Phase One was learning phase: Phase Two was eye movement experiments; Debriefing sessions were conducted in Phase Three.

In Phase Two, participants were asked to perform browsing on each Web page in their native language version. For the browsing task subjects were asked to look through the Web page for thirty seconds, and asked the following questions to motivate their browsing before the subject see the Web page: “You are going to browse a Web page about 30 seconds. After you brow a Web page, you will be asked about the Web page’s visual presentation (good and bad). How do you like the Web page in terms of visual presentation?”

The subjects were given about 30 seconds for browsing the Web page and eye tracking systems recorded eye movement data until the Web browser was closed by the experiment administrator.

3 Data Analysis and Results

Although the subjects were given about 30 seconds for browsing tasks, only first 15 seconds of eye movement data were analyzed. ClearView 2.7.1. Software provided with the eye tracker helps to handle the volume of data and analyze the data. To have accurate data, a validity and eye filter were used. By reviewing the validity code[1], only one American participant should be excluded from the entire data. The data from this participant had many missing sequences of gaze coordinate.

3.1 Browsing Task on the Simple Web Page

For the browsing task on a simple Web page, the LexisNexis global Web page was used. As shown in Fig. 1, the Web page was divided by eye the tracking software into seven AOIs, which are invisible to the participants. The Korean Web page for this task also had the same number, almost identical in size and location of the AOIs.

American participants fixated and allocated their fixation on the banner image (*faces*) and banner text area (*btext*) more often than Korean participants: American participants allocated their fixations about 31% on the banner image (*faces*) and the banner text area (*btext*), while Korean participants gave about 25% of their fixations on those areas. Korean participants gave more attention to the right top area (*righttop*) and learn about area than American participants. It is interesting that although the banner text is expected to be viewed by all participants about 26 % of Korean participants (5 participants out of 19 participants) did not fixate at all on the banner text in the 15 second frame. More than half of the participants in each group did not fixate on the bottom area of the Web page. Although there are three images (logo, map, and banner image) on the Lexis-Nexis Web page, participants tended to fixate

[1] The validity code is a measure of the system’s certainty that it has recorded the correct data ranging from 0 to 4, and is used for data filtering. The manual recommends removing all data points with a code of 2 or higher: 2) recorded one eye, and has no way of determining if this is the left or the right eye, 3) the actual gaze is incorrect or corrupted, and 4) the actual gaze data is missing or definitely incorrect (Clearveiew 2.7.1 user manual, p. 17).

Fig. 1. LexisNexi Web page for American

less on the images the except banner image than text areas such as login and learn areas on the left column.

The Korean participants spent about 49 percent of their total gaze time on three AOIS: *learn* (21%); *empty space* (14.7%); and the *faces* AOI (13.1%). American participants also allocated about 47 % of their gaze time on the same AOI and empty spaces as Korean participants did. The average gaze time on the banner image (bimage) and the banner text (btext) for American participants is higher than that of Korean participants. However, an average gaze time on learn about area (learn) for Korean participants is higher than that of American participants. Corresponding to the fixation count and average gaze time, Korean participants gave less attention to the banner image (faces) and the banner text (btext) than American participants, but allocated more fixation count and gaze time in viewing the learn about area (learn), which is written in English, than American participants.

Both groups of participants fixated first on the *faces* (average time to first fixation for face :AME 530msec and KOR: 1136 msec), followed by *logo* (average time to first fixation for logo:AME: 2351 msec, KOR:1655msec). Almost 80 % of the Korean subjects and 75 % of American subjects fixated first on the banner face image, and then about 80 % of them fixated on the logo area. Interestingly, Korean participants reached the banner text area relatively late.

The Korean participants fixated on the right column area of the Web page such as map, login, and learn AOI faster than the American subjects did, while the American participants gave more attention and spent more time on the banner image (*faces*) and the banner text area (*btext*) than the Korean participants. However, there was no difference in the fixation transition among the AOIs.

3.2 Browsing Task on the Medium Complex Web Page

For the browsing task on a medium complex Web page, the Cisco Web page was used. As shown in Fig 2, the Web page was divided into 11 AOIs.

The right navigation area (*rnavi*) that has 12 links received the most fixations with an average of about eight fixations and about nine fixations from the American participants and the Korean participants, respectively. Both groups of participants allocated about 30% of their fixations on the banner image area (*bimage*) and the banner text area (*btext*) in the center of the Web page, and about 36 % on the login area (*topr*) in the upper right of the page. However, latest news area (*news*) and featured product area (product) received an average of 3-4 fixations from the subjects for both groups. American subjects fixated more on banner image than banner text, while Korean subjects gave more attention to the banner text area (*btext*) than the banner image area (*bimage*). Interestingly, two small images (*twoim*), which were located right below the banner image, did not receive any fixations from the American participants at all, and received only one fixation from a Korean participant. Apparently, the log area (*logo*) in the upper left corner is the area that received the least fixations on average along with the bottom area (*bottom*).

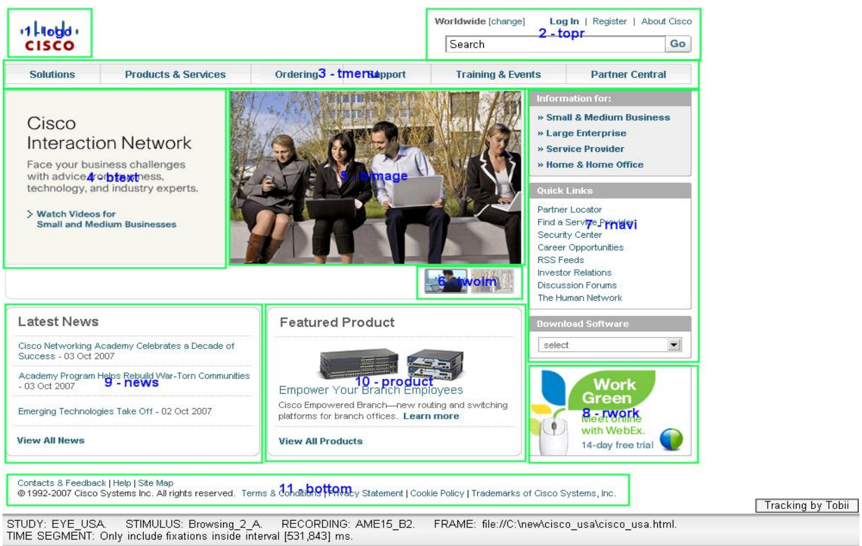


Fig. 2. Cisco Web page for American

The American participants and the Korean participants spent almost the same percentage of gaze time (AME: 49.1%, KOR: 49%) on the upper area of the home page such as, the logo area (*logo*), the top right area (*topR*), the top menu (*tmenu*), the banner text area (*btext*), and the banner image area (*bimage*). The Korean participants spent more time on the banner text area (*btext*) and the right navigation area (*rnavi*) than the American participants, while the American participants spent more on the top right area (*topr*), and banner image area than the Korean participants. The Korean participants allocate their gaze more on the banner text area (18.4%) than banner image area (9.5%), while the American participants spent slightly more on the banner image area (14.3%) than the banner text area (12.4%). Other areas of the Web page such as logo, two small images under the banner images, news and product areas received similar gaze time and allocation of gaze time between two groups.

Sixteen participants for both groups, about 88% of American and 84% of Korean participants, fixated first on the banner image. About 45% of American participants moved to the banner text area as a second target area followed by the top menu, whereas the Korean participants' second target was the top menu, logo, or banner text area with about 31% each.

Overall, the AOIs receiving the earliest fixation from the participants for both groups were center and top left areas such as, banner image area (*bimage*) and banner text area (*btext*), logo area (*logo*), and top menu area (*tmenu*). Areas under the banner image received their first fixation relatively later than upper parts of the Web page.

American subjects fixated more on banner image than banner text, while Korean subjects gave more attention to the banner text area (*btext*) than the banner image area (*bimage*). The Korean participants spent more time on the banner text area (*btext*) and the right navigation area (*rnavi*) than the American participants, while the American participants spent more on the top right area (*topr*), and banner image area than the Korean participants. Common gaze sequences for both groups followed clockwise or counter clockwise directions.

3.3 Browsing Task on the Complex Web Page

For the browsing task on a complex Web page, the Seagate Web page was used. As shown in Fig 3, the Web page was divided into 14 AOIs.

The left navigation area (*lnavi*) received the most attention for both groups of participants (AME: 189, KOR: 184) with an average of about 11 fixations, followed by the banner text (*btext*) for American participants and the top customer story (*cstory1*) for Korean participants. The American participants gave more attention to the banner image (*bimage*) and the banner text (*btext*) areas than the Korean participants, while the Korean participants allocated their attention more on the upper customer story area (*cstory1*) and right story area (*rstory*). However, both groups did not read the whole story, while some participants seemed to read the whole story. The rest of the AOIs did not receive a significant number of fixations. The left navigation area (*lnavi*), the right story area (*rstory*), and the upper customer story area (*cstory1*) received good attention from both groups of participants.

The Korean participants spent slightly more time on the content areas such as the right story area (rstory) and the first customer story (cstory1) than the American participants. The American participants allocated their gaze time more on the banner image and the banner text area than the Korean subjects. Participants for both groups spent about 65% of their gaze time in viewing the banner image and the banner text, the left navigation area, and the right column. Participants allocated less than 20% of their gaze time for viewing the content displayed in the center column. They spent only 12% of their gaze time in viewing administrative content such as the logo, the account, the country selection, and the foot menu.

The American participants made the most frequent transitions from the other AOIs to the banner image, while the Korean participants made the most frequent transition from the AOIs to the text contents area.

Overall, participants for both groups seem to have similar patterns of eye movement in the browsing tasks. Participants for both groups tend to see upper section of the Web pages more often and relatively long. The level of Web complexity does not seem to affect participants' initial attention. Korean participants tend to spend more time on navigation area than American participants, while American participants spent more in viewing banner images than Korean participants. With respect to scanpath in browsing tasks, common scanpath were not found. However, more various scanpaths were found in the complex Web page.

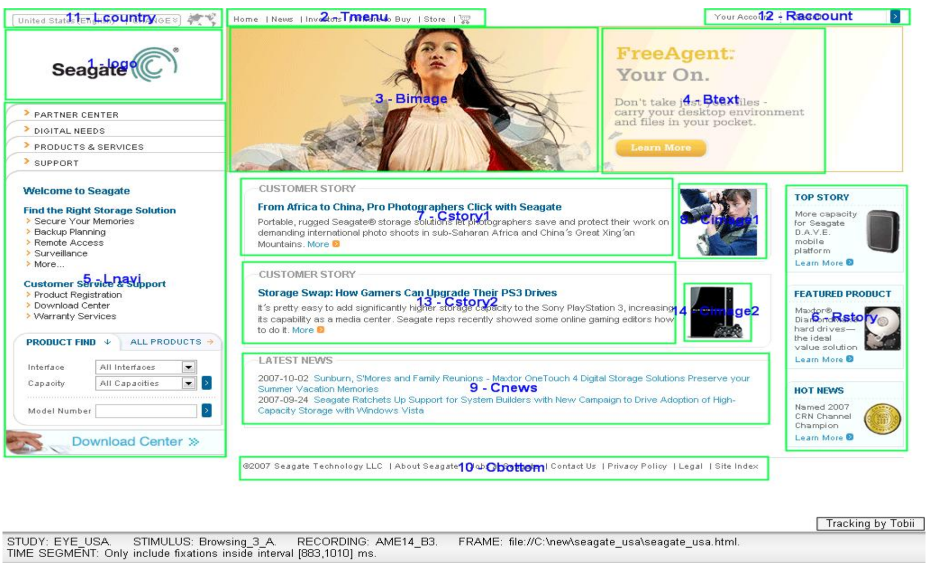


Fig. 3. Seagate Web page for Americans

3.4 Cultural Differences in Attention and Viewing Pattern

The differences in terms of allocation of viewing time between American participants and the Korean participants were not found except that American participants spent more time in viewing banner images than Korean participants, while Korean participants spent more time in viewing navigation or menu areas. In the current study there is no significant difference between American participants and Korean participants in terms of initial attention to visual elements and areas on the Web pages.

However, there are differences in the allocation of fixation and gaze time on the AOIs, especially on the banner images and navigation areas, in browsing tasks. American participants tended to give more attention to the banner images than Korean participants in browsing tasks on all levels of complexity of Web page. Korean participants tended to allocate their fixations and spend more time in viewing navigation areas. This finding somewhat supports Nisbett et al.'s [7] holistic versus analytic cognition and partially supports Chua et al.'s [8] study.

Although this study was investigating similarities and differences in viewing patterns between both groups, it is difficult to show the differences and similarities with statistical evidence due to the lack of tested methodologies that compare scanpaths between groups. According to Chua et al.'s findings, the Korean participants who have a holistic view point would be assumed to reach the AOIs sooner than the American participants. However, there is no significant difference between the two groups. In browsing tasks on the medium complex and complex Web pages, the numbers of AOIs visited over a 21 second period is similar.

4 Future Research and Conclusion

Due to the small sample size differences between groups might not be observed in this study. For future research, it is necessary to examine ocular behavior when viewing a Web page with a large sample size to provide statistical evidence in various types of Web pages with many different tasks. It will also be very interesting to see how increasing cognitive loads affects ocular behavior in different cultures when users search, or view, a Web page. Since many Web pages include multimedia elements such as video or audio, it is also necessary to examine how ocular behavior is affected when users interact with a multimedia Web page. The data obtained from this study reveal there are similarities existing in terms of attention and viewing patterns among individual participants. This also proposes future research questions. It is necessary to examine whether correlations exist between individual differences and eye movement characteristics. In addition, it should be explored which individual factors affect more on eye movement than cultural factors.

The fundamental research questions for this study are to understand whether there are cultural differences in ocular behavior when viewing a Web page in order to provide practical insight and guidelines for designing Web pages for a particular cultural group. Although this study could not provide strong evidence that there are cultural differences in eye movement behavior when viewing a Web page, this study shows many cultural variations in eye movement when viewing a Web page with different tasks and raises methodological issues in eye movement research for cultural studies in the files of information science.

Acknowledgements. The eye tracking systems were set up and experimental data were collected in the usability lab in College of information at Florida State University from June 2, 2008 through July 2, 2008.

References

1. Badre, A.: The effect of cross cultural interface design orientation of World Wide Web user performance, `http://www.cc.gatech.edu/gvu/reports/2001/abstracts/01-03.html`

2. Callahan, E.: Interface design and culture. In: Cronin, B. (ed.) Annual Review of Information Science and Technology, vol. 39, pp. 257–310. Information Today Inc., NJ (2005)
3. Cyr, D., Trevor-Smith, H.: Localization of Web design: an empirical comparison of German, Japanese, and United States Web site characteristics. Journal of American Society for Information Science and Technology 55(13), 1119–1208 (2004)
4. Marcus, A., Gould, E.W.: Crosscurrents: Cultural dimensions and global web user interface design. Interactions 7(4), 32–46 (2000)
5. Singh, N.: Analyzing the Web as a cultural document: an exploratory study of cultural values as depicted on country-specific Web sites. Unpublished Dissertation, Saint Louis University (2003)
6. Faiola, A., Matei, S.: Cultural cognitive style and web design: Beyond a behavioral inquiry of computer-mediated communication. Journal of Computer-Mediated Communication 11(1), article 18 (2005), http://www.jcmc.indiana.edu/vol11/issue1/faiola.html
7. Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and systems of thought: Historic vs. analytic cognition. Psychological Review 108, 291–310 (2001)
8. Chua, H.F., Boland, J.E., Nisbett, R.E.: Cultural variation in eye movements during scene perception. PNAS 102(35), 12629–12633 (2005)

Trails–An Interactive Web History Visualization and Tagging Tool

Wenhui Yu and Todd Ingalls

School of Art, Media and Engineering,
Arizona State University
Tempe, AZ 85287
{Wenhui.Yu,Todd.Ingalls}@asu.edu

Abstract. In this paper, we described an innovative web history visualization and tagging tool – Trails, which is designed and developed to help people understand their browsing history and habits better. We gathered users' impressions of using Trails, including comparison with traditional web history views, perceived usefulness, privacy concerns, and suggestions to improve the system.

Keywords: Web browsing history, peripheral awareness, information visualization, personal informatics, on-line activities.

1 Introduction

The World Wide Web is said to be the most popular global communication medium and knowledge repository, which is now the platform for a myriad of activities [1]. As more and more human activities have come to rely on the use of the web, people are spending significant amount of time everyday interacting with and through web browser software, to search, work, entertainment and communicate [3]. Thus, our web browsing history is increasingly reflective of our interests, needs and what we do in our daily lives [3]. A class of systems and research about personal informatics has shown that people strive to obtain self-knowledge, collect and reflect on personal information [2]. From this aspect, exploring people's web browsing history is worthy as an interesting direction of study in personal informatics, which is potentially valuable to track our life, understand our habits and foster our awareness [3].

Moreover, people's problematic Internet use problems caused by overwhelming information provided on the Internet also motivate us to explore people's web browsing history and look for solutions. For example, modern office life has shown an increasingly common condition called "attention deficit trait" because of information overload, which might cause productivity losses [4]. A myriad of research work [5][6][7] related has been done around the problematic Internet use behaviors like severe procrastination, Internet addiction and compulsive Internet use, which might result in further problems of stress, depression and sleep disturbances [7][8]. However, people's limited memory determines that they cannot observe some behaviors directly and may not have time to constantly and consistently observe some behaviors [2]. Thus, people are not all capable of discerning what time being spent on

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 77–86, 2011.

the web is helping us get things done or distracting us from things we want to do. In orders to remind people of these problems, tools to help us reflect on our web activities are needed.

Web history storage and retrieval services, provided by most popular browsers, have the potential to serve as a good resource to track our on-line visits. However, the current web history viewers in most popular browsers lack the way to help people reflect on their web activities. This paper mainly introduced an interactive web history visualization and tagging tool, Trails, which was designed and developed to visualize web history data beautifully and informatively in the browser. Through the aid of visual elements, users are expected to retrace their life on Internet and understand their habits better. It is also expected to help users to search and organize the web content faster, easier and more fun.

2 Problems, User Needs and Design Goals

In order to narrow the problem scope and understand user needs, we did user survey among 10 browser-heavy users whose life largely depends on computer and Internet before we developed system. In the survey, other than some questions to understand their self-report browsing habits, we asked them to look into their own web history in a time period and asked them some questions like: “What is your main activity in that time period?”, “Other than the main activity, what other activities you did?”, etc.. And time they spent on figuring out answers was recorded for future comparison with our system. This is very helpful for us to figure out limitations with existing implementations [6]. Take Firefox for example:

- It is not efficient for people to track how frequent they visit a certain URL in a specific time period. The history viewer only shows visits for a whole day in a table, which means users had to locate the time and parse information in that time period by scrolling the table;
- There is a lack of summarization methods to allow people to investigate their own browsing habits. In order to get accurate answers, users have to conclude their behaviors by looking at items in the table one by one;
- There’s no effective way for people to categorize their history visits. Bookmarks are used for storing information of interest, but useless to track the whole browsing trails.

We should address these problems in our work. Moreover, in order to provide better user experience, we tried to ask participants to give feedbacks on solutions provided by some time management softwares (e.g. ManicTime[1]) and other web history visualization tools (e.g. Slife[2]). Then we determined our design goals based on the analysis of results:

- **Intuitive.** Information visualization has been proposed as a way to cope with the problem of “lost” by taking advantage of people’s innate perceptual skills to support cognitive skills [1]. Thus, much like other web history visualization tool,

---

[1] http://www.manictime.com/

[2] http://www.slifeweb.com/

such as Eyebrowse [3], our tool should produce easy-to-read statistical summary visualization for time-management reminder purpose.

- **Unobtrusive.** We brought up this goal when we asked some feedback on Parental Control. Participants indicated that it is an obtrusive way to change people's browsing habits. Based on this consideration, we would like to provide an unobtrusive way to foster awareness instead of annoying people by alarm-like notification or surveillance.
- **Minimal Cognitive Load.** By showing some time-management tools to users, like ManicTime, we realized complicated UI and visuals might be confusing. And majority of participants indicated that they would not like to spend much time on this kind of tool, although they are curious about keeping informed of their on-line activities. Thus, a solution is needed for our tool to help users informed with least mental efforts.

3 Trails

We developed Trails as a Firefox add-on. After installed, one icon will be added at the upper-left corner of UI, which can trigger the Trails application with full features and functions in a new tab view in the browser.

In this section, this paper describes Trails mainly from two aspects: features that make it distinctive, design and implementation details.

3.1 Features

There following are novel features of Trails:

- Peripheral awareness

From the survey, we figured out that people are not willing to spend much time or take efforts on figuring out exact information about their browsing history, while most people are only interested get a general understanding within least time and efforts. Thus, Trails provides users an instantly understandable summary of their activities using several key metrics of visual element (size, shape, color). This strategy utilizes peripheral awareness concept, for the purpose of helping people quickly access, interpret and keep awareness of activities, while at the same time avoiding needless distraction from their main tasks [9].

- Hour-based data retrieval

Trails provides an hour-based view during any given day, so users can easily locate a specific hour or a period of time in a day to understand their activities instead of reading a table. And Trails's timeline was also designed as horizontally aligned "hour blocks" to help users to compare their activeness during a day. This will be elaborated in the design details part.

- Tagging

Tags are not often used in other browsers' built-in web history viewers. Firefox has tagging service, but rarely used according to our survey. Compared to more common service - bookmark, tagging can provide better and more detailed inspections on whole browsing preferences and life habits. Trails differs from similar efforts by providing an innovative method to organize, tag and display information in intuitive ways, so that people can retrace their web visits based on tags through limited glances.

- Animated representation and interaction

Through the aid of visual element and animation effect, users can search and tag the web content faster, easier and more fun. Meanwhile, animations and transitions are very necessary for users to build mental maps of spatial information [10]. This feature is especially useful in filtering content, toggling among views and comparing among different time periods.

3.2 Design and Implementation Details

In this section, the paper will explain the design details of Trails from visual representation and user interaction perspectives, and explain how the design decisions were made from consideration of needs and goals.

Visual Representation. Trails mainly have 4 views: Individual View, Group View, Sort View and Summary View. Every view has its own basic visual element. Different visual elements move, gather together or disperse to build views to facilitate people's understanding.

Visual Elements

Figure 1 (a) describes the basic visual element in Individual View and Sort View, which is the favicon of the visited web page with a tagging ring, representing a URL. The size of the ring is decided by the visit count of the URL in the selected time period. The more visits on an URL, the bigger the ring. The tagging ring can divide into several parts according to the tags the user associates with an URL. Every color represents a corresponding tag. If there's no tags associated, the ring will present with a grey color. For example, in Figure 1(a), there're 4 different tags associated with the URL, which represents 4 different activities assigned on the ring by the user.

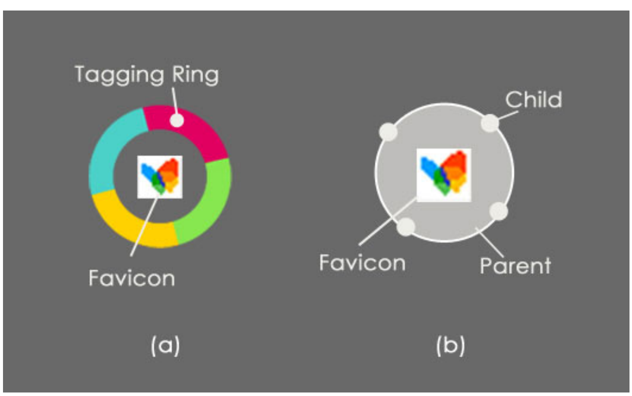


Fig. 1. The visual elements used in different views. (a) is the basic visual element for Individual View and Sort View. (b) is the basic visual element used in Group view.

Figure 1 (b) is the basic visual element used in Group View, which represent a group of URLs with the same domain instead of an individual URL. The favicon of the domain is positioned in the middle (It is possible that URLs within same domain using different favicons). And the size of parent is determined by the total visit count in this group. Children attached on the parent are member URLs in this group, which distribute evenly on the edge of the parent.

Views. There are four views designed, each with different representation emphasis. These four views share same UI controls like search bar, date picker, tags viewer and timeline. Shown in Figure 2, 3, 4, 5, Trails's timeline is 24 horizontally aligned "hour blocks" with different transparency proportional to the total visits count during that hour. We hope users can understand their activeness of a day just by looking at the timeline.

- Individual View

In the Individual View, Trails loads all the visited URLs in the specific hours and displays them as tagging rings, which spread around the view and move slowly along with time. The connection link between different basic elements shows their "from-to" relationship. Individual View is designed for general understanding of users' browsing activities. Users can tell the top visits from the size, the attributes of the visits from the colors and the path of visits from connection links with little mental effort.

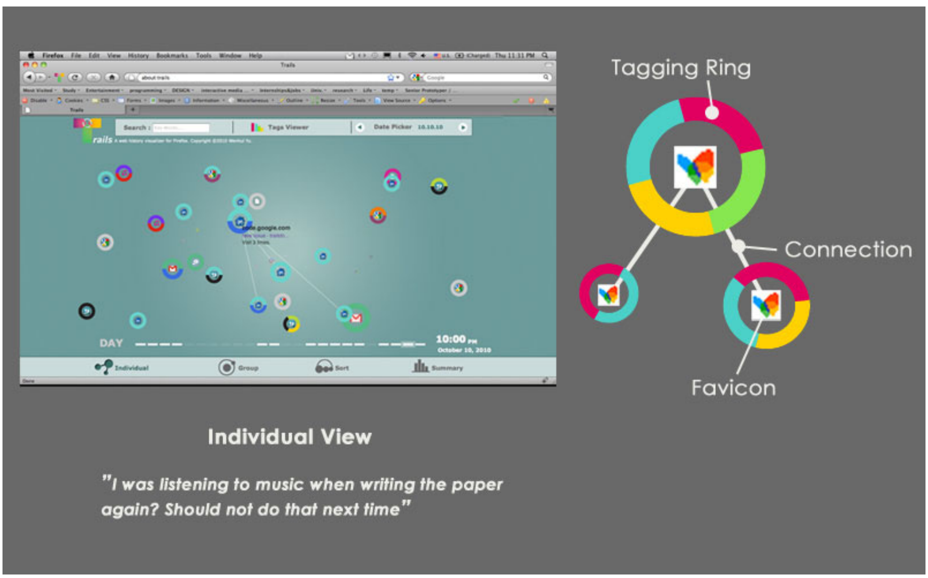


Fig. 2. Individual View

- Group View

Group View is designed for understanding activities in a certain site rather than an individual URL. As mentioned above, URLs within the same domain will be grouped together as a parent circle with child dots on it (Figure 3 (a)). (a) can be expanded into an orbit-like system by clicking, showing details representation of the group (Figure 3 (b)). And when user click inside the (b), it will collapse back to parent-child look. This solution provides two level of information: by looking at the parent-child visual, users can understand their browsing habits on certain sites in general; by looking at orbit-like visual, users can understand details on their activities in the site.

In Figure 3 (b), the positions of URL tagging rings on the orbits in the same group are decided by the nodes' level in the domain tree [11]. Tagging rings on the same level will be put into the same orbit; the lower the level, the further the orbit from the center. Thus, users can understand their activities and habits in a certain site by looking into the color distribution and shape of orbits instantly, instead of parsing every individual URL. For example, if a user browsed his friend's new album with 20 pictures on Facebook.com. The orbit-like visual will probably show a large orbit with 20 tagging rings distributed evenly on it, which is very easy to cognize because of its distinctive appearance. And if there're tags associated on the tagging rings, that will provide more information to help people understand.

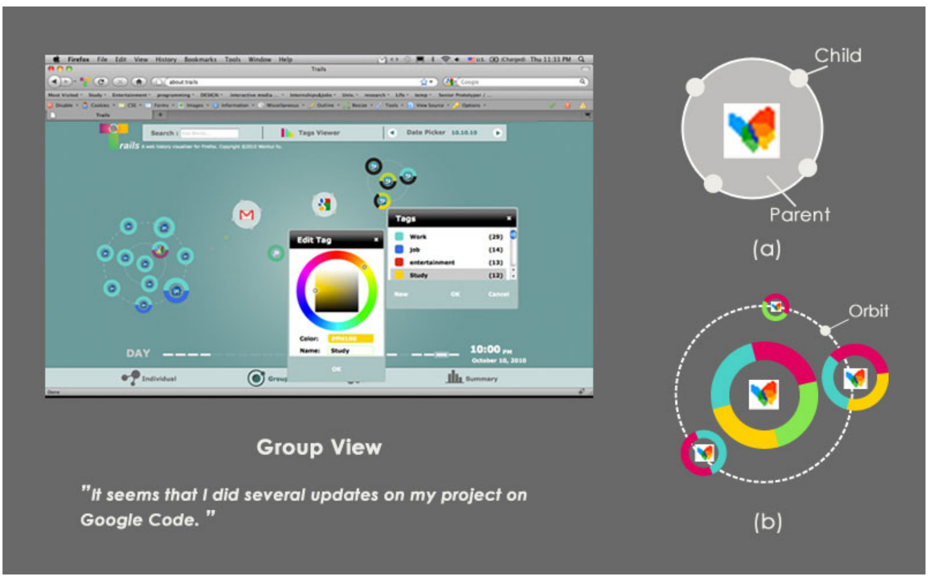


Fig. 3. Group View

- Sort View

In Sort View, URL tagging rings are ranked by visit count, so that users will have good understanding about their browsing habits. When the user trigger the Sort View button, the originally freely moving URL tagging rings will move into a horizontal line from left to right. This view is designed to foster people's awareness of their browsing frequency on certain sites, which is especially for people who have compulsion problem. For example, some people might check their Facebook profile page or email account every 5 minutes, even when they know there probably won't some useful news for them.

People can also navigate to figure out their visit paths in Sort View, as in the Individual View. However, different from Individual View, connections in Sort View are drawn in arc, instead of line, so that it's easier for people to read and comprehend.

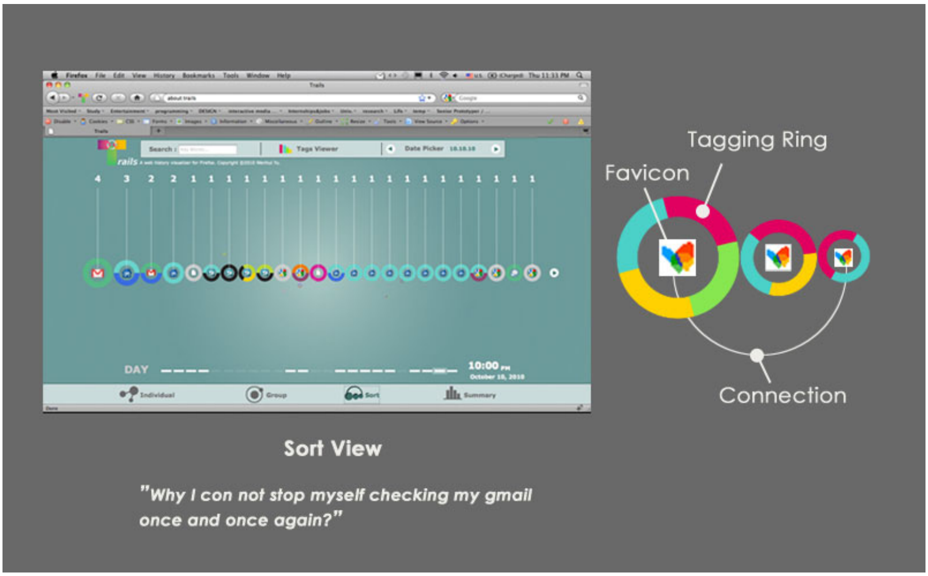


Fig. 4. Sort View

- Summary View

Summary View displays visit information based on the tags associated. On the left of view, there's a wheel-like chart showing popularity of tags. In the example of Figure 5, right part of view displays details of all URLs associated with the " lifestyle" tag, with total visits information and visit count for every URL. Summary View is designed for giving users an intuitive understanding of their on-line activities.

User Interaction. Other than visual representation, user interaction is another important aspect of design.

Navigation. Compared to traditional Firefox web history table view, navigation can be a challenge for the current design because of its unordered attribute. Thus, introducing search bar is very necessary to help user navigate in the browsing history. In Individual View (if the current view is not Individual View when search bar activated, the application will jump to Individual View), User can search URLs by typing into the search bar at the top-left of Trails' UI. During User's typing, the URL tagging rings unrelated will disperse and move outside of the view.

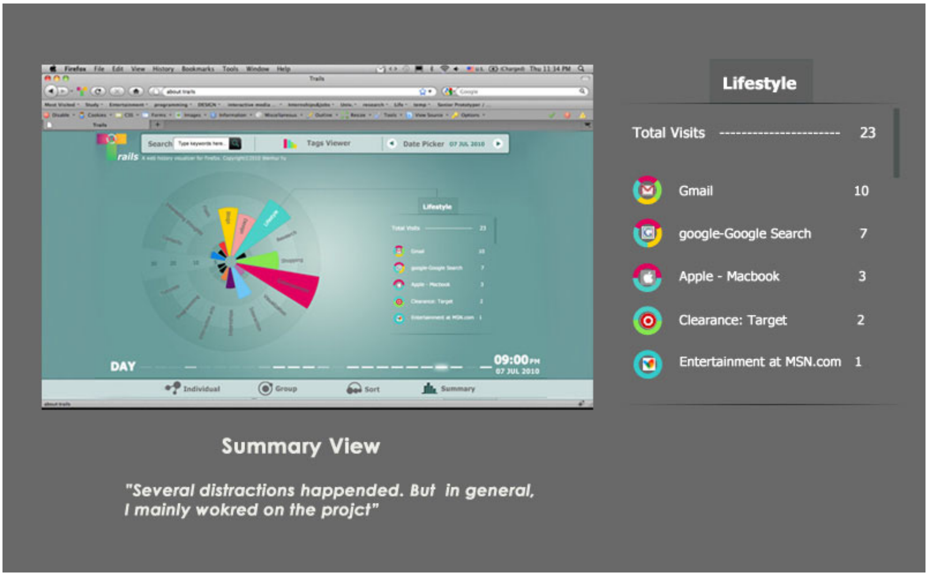


Fig. 5. Summary View

Tagging and un-tagging. As one of the most important feature of design, users have several options to tag or un-tag the URL easily and fast.

- Right click on tagging rings can remove the tag instantly;
- Double-clicking on the tagging ring will trigger the tagging panel, where users can attach tags, remove tags or create new tags.

Locate to a specific time period. In order to locate to a specific time period, user should open the date picker on the top right to pick a date. Then users can inspect specific hours' activities by clicking on those hour blocks. When the user is clicking on the block, the URL tagging rings in that hour will move into the view; click again, the URL tagging rings will disperse and move outside of the view.

4 Initial Study

In order to explore usefulness of Trails, initial user study has been conducted among 5 graduate students selected from the preliminary survey. Our study focused on comparing the effectiveness of using Trails to understand users' browsing activities with using traditional Firefox web history viewer, using same period of well-tagged visits, during a limited time, by few glances. The study has two sections:

In the first section, users were asked to understanding the visuals in the 4 views. Users were asked the question same in the preliminary survey, and the time they spent on figuring out answers were recorded. 5/5 participants reported they found Trails is more helpful to help them to understand their browsing habits and activities compared to the traditional table view. 4/5 participants can make successful conclusions by using first person narratives on browsing habits (see quotes in figure 2, 3, 4, 5). One participant even can tell that "I slept very late" just by looking at hour blocks.

In the second section, users were assigned some tasks to complete. These tasks have: toggling among views, searching a specific content, creating new tag, tagging and un-tagging. Questions related to the participant's impressions about using Trails were asked. 3/5 participants reported interacting with tagging circles is more fun compared to selecting from tagging list, but not very efficient.

5 Conclusion and Future Work

For the next step, we will install Trails in users browsers and let them try it for two weeks for further in-depth feedback. But based on our initial study, although most participants think Trails is easier to understand their browsing habits and provides a way to organize and categorize their activities, some individuals proposed some concerns on usefulness of Trails, and we also figured out some problems when observing users' using, which motivate us to go further on exploring better solutions.

For example, we realized the need to tag the URL without navigating views. Several participants complained that they didn't have time to play with those rings to tag. Thus, we added a small button at the bottom right corner of browser to trigger a small tagging panel for quick tagging a web page without launching Trails. We expect to test the effectiveness of the in the future study.

A participant indicated that it is better to automatically tag some popular URLs or sites instead of asking users to tag everything; another user commented that there should be several ways to group URLs. Thus, we plan to explore the more complicated methods to categorize and group URLs, and correlate these methods with creative visual presentation to facilitate reflection.

Finally, we think Trails is a beneficial attempt, although a lot of future work is needed. For example, all uses showed great interest to explore Trails and were excited about what Trails displayed at the first time, but some people thought they would not use Trails seriously because of lacking motivation to do self-reflection. Thus, we would like to explore better approaches to motivate people to use it. Eyebrowse [3] is a good example by introducing real-time web activity sharing in a social platform, which shows us another way to investigate in the future.

References

1. Turetken, O., Sharda, R.: Visualization of Web Spaces: State of the Art and Future Directions. ACM SIGMIS Database 38(3) (August 2007)
2. Li, I., Dey, A., Forlizzi, J.: A Stage-Based Model of Personal Informatics Systems. In: CHI 2010, pp. 557–566 (2010)
3. Eyebrowse: Real-Time Web Activity Sharing and Visualization
4. Hallowell, E.M.: Overloaded circuits: Why smart people underperform. Harvard Business Review, 55–62 (January 2005)
5. Young, Case, C.J.: Internet abuse in the workplace: new trends in risk management. Cyberpsychology and Behavior 7(1), 105–111 (2004)

6. Treuer, T., Fabian, Z., Furedi, J.: Internet addiction associated with features of impulse control disorder: is it a real psychiatric disorder? Journal of Affective Disorders 66(2-3), 283 (2001)
7. Whang, L.S., Lee, S., Chang, J.: Internet over-users' psychological profiles: a behavior sampling analysis on internet addiction. Cyberpsychology and Behavior 6(2), 143–150 (2003)
8. Thomee, S., Eklof, M., Gustafsson, E., Nilsson, R., Hagberg, M.: Prevalence of perceived stress, symptoms of depression and sleep disturbances in relation to information and communication technology (ICT) use among young adults - an explorative prospective study. Computer in Human Behavior 23(3), 1300–1321 (2007)
9. Cadiz, J., Czerwinski, M., McCrickard, S., Stasko, J.: Providing elegant peripheral awareness. In: CHI 2003 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, April 05-10 (2003)
10. Bederson, B., Boltman, A.: Does Animation Help Users Build Mental Maps of Spatial Information. In: Proceedings of InfoViz 1999. IEEE, Los Alamitos (1999)
11. Gandhi, R., Kumar, G., Bederson, B., Shneiderman, B.: Domain Name Based Visualization of Web Histories in a Zoomable User Interface. In: Proceedings of the 11th International Workshop on Database and Expert Systems Applications, September 06-08, p. 591 (2000)

Listen! Somebody Is Walking towards Your Car (Introducing the Awareness-3D Sound System into the Driver to Increase the Pedestrian's Safety)

Mohammad Ardavan and Fang Chen

Interaction Design Group, Department of Computer Science
Chalmers University of Technology, Gothenburg, Sweden
ardavan@student.chalmers.se, fanch@cs.chalmers.se

Abstract. Car accident statistics indicate that the pedestrians are the majority of the road traffic victims due to drivers' lack of adequate visibility on the road. In this paper, the effects of human natural sounds in drivers' awareness were investigated in order to increase pedestrian safety by carrying out a study on introducing a 3D sound system into drivers. All studies with collected results showed strong positive support to the design of using 3D sound system to present the pedestrians' situation to car drivers.

Keywords: Safety, pedestrian, car driver, sound, 3D sound.

1 Introduction

One of the most important global wised problems in traffic safety is pedestrian safety in urban areas. Traffic accidents were described as one the main causes of death and injuries around the world by a World Health Organization. They report an estimated 1.2 million fatalities and 50 million injuries every year, so the safety issues are catching more and more attentions, therefore many studies focus on how to improve drives by effectively educating the driver to have safe driving, by designing better Human-Machine Interface and informing drivers about when and where to put on attention. At the same time active safety system are developed for the intention of avoiding accident happening ([1]; [2]). In the other hand, car accident statistics indicate that the pedestrians are the majority of the road traffic victims due to drivers' lack of adequate visibility on the road. In this paper, the effects of increasing drivers' traffic awareness by using 3D natural sounds to indicate the location of pedestrian to the driver were investigated in order to increase pedestrian safety on the road. Eyes and ears are the two most import sensors for human beings to perceive information. To have safety drive, the drivers' eyes should keep on the road as much as he can. At the same time, there are already many different kinds of drive directly and indirectly related information that presented to the driver visually nowadays. While inside the car, it is relative quiet with good noise isolation. To reduce the visual information overload, human's audio information perception needs to be explored. Therefore, the research question of the study is "Do the 3D natural sounds have possibility and usage to present the pedestrian's location and moving direction to driver?" The possibility

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 89–98, 2011.

means to investigate whether a 3D natural sound could simulate the pedestrian position for the driver or could be a cause of destruction and discomfort to the driver. The usage means to inspect whether a 3D natural sound could prevent the immediate threaten and guide the driver's attention towards the right direction. Aviation study showed that using 3D audio signal could reduce the target searching time therefore increase the pilot's situational awareness ([5]; [4]). Tan and Lerners' study indicates that directional acoustic cues to hazardous direction have the potential to speed up driver response [6]. Chen has carried out a couple related study [2] and show that 3D natural sounds can be used for presenting traffic information in order to provide the traffic situation of all road users to truck driver.

2 Methods

To have a better understanding of traffic situation on the road, the first step in this study was to investigate the video records in the real world. The video cameras for traffic monitoring were placed in different crowded locations in Dalian, a city located in the north of China and in Gothenburg-Sweden. Through the investigation of the videos recorded, the pedestrians were categorized into six types: 1. Man/ Woman, 2. Old man/Old woman, 3. Pregnant woman/ Woman with baby carriage, 4. Child, 5. Handicap, 6. Crowd.

Designing the proper natural sounds was the main challenge of this study. Different natural sounds were selected in order to present the six types of pedestrian to the driver. Steps sound was selected as base sound. For having more distinguishable natural sounds, some specific human's sounds were added to steps sounds like old man's coughing sound, baby's crying sound. In order to improving the quality of sounds, the format of sounds were changed to WAVE and the noise was removed, then 48000 Hz and 32-bit were selected as their frequency and bit ratio according to hearing ability of human. To make the directional sounds, Fade in and Fade out effects and switching method between two speakers were used. All the levels were done by using Audacity 1.3 software.

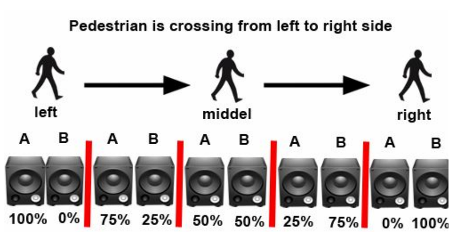


Fig. 1. Directional Sound - Switching method between two speakers in exact time

To examine how 3D sound can help the drivers and draw their attention into moving direction of pedestrians, Simulation study was carried out in the lab using STISIM. Eight loud speakers were placed as square-shaped in four different directions (forward, back, right and left) around driver in which the driver's seat was at the center. Fifteen students, twelve male and three female, age 21-33 years, were randomly selected as subjects to this study. All of them have driving license.

The lab study is divided into three parts:

First part: the representative of the sounds (S)
In this test, all the users listened to eight different natural sounds that represent the 8 types of pedestrian. The purpose is to understand the meaning that carried within the sound. After each sound, some questions were given to collect their understanding of the sound. The sounds were S1: man, S2: more than one man, S3: Old man, S4: Child, S5: Children, S6: Crowd, S7: Pregnant woman, S8: Handicap.

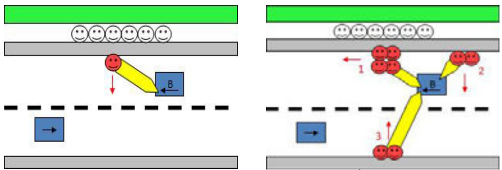
Second part: the direction of the sounds (DS – Directional Sound)
In this test, all the users listened to ten directional sounds. After each sound presentation, some questions were given for direction perception. The sounds were: DS1: Old man sound from forward, left to right, DS2: Pregnant Woman from back, right to left, DS3: Man sound from right, back to forward, DS4: Old man and man sounds from forward, left to right, DS5: Children sound from right, back to forward, DS6: Old man sound from forward, left to right and Pregnant woman sound from right, back to forward, DS7: Child and Old man sounds from forward, left to right and Man sound from forward, right to left, DS8: Old man sound from forward, left to right, Man sound from right, back to forward and Pregnant woman sound from back, right to left , DS9: Man sound from forward, right to left, Man sound from forward, left to right and Child sound from right, back to forward, DS10: Crowd sound from all directions

Third part: implementing 3D sounds to drive simulator
To have deeper understanding of how 3D sound can help the driver to detect the pedestrian in different situation, two sets of six simulation scenarios were developed in STISIM about car and pedestrian's interactions in urban areas. The purpose is test if the driver can perceive pedestrian's location and moving direction correctly by 3D natural sounds presentation. Each scenario was organized in order to represent one potential dangerous situation. After each scenario, some questions were given to the subjects. The questions included the questions about 1) if 3D sounds help the subjects distinguish the direction of specific pedestrian's movement; 2) if 3D sounds act like alarm to draw their attention to dangerous situation; 3) if car's engine sound can be a cause of problem in hearing the 3D sounds; 4) if car's audio system can be a cause of problem in hearing the 3D sounds. All subjects driving procedure was recorded in order to measure the drivers reactions like driver's immediate threaten or in time threaten or inattention in different dangerous situations.

2.1 First Set of Scenarios

1-1st Scenario. Driver B is informed that one man is crossing the road from crowded area from his front side, moving from right to left using S1 sound.

1-2nd Scenario. Driver B is informed that the children (no.1) from his right side moving from back to forward using S5 sound, one man and pregnant woman (no.3) from his front side moving from left to right using S1 and S7 sounds and one man and old man (no.2) from his back side moving from right to left using S1 and S3 sounds are crossing the road. All pedestrians are represented at the same time to check if driver has enough ability to distinguish five different directional sounds at same time or not.

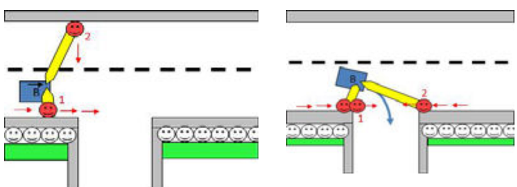


1-1st Scenario 1-2nd Scenario

Fig. 2.

1-3rd Scenario. Driver B is informed that one man (no.1) from his right side moving from back to forward using S1 and one child (no.2) from his front side moving from left to right using S4 are crossing the road. The two sounds play together at same time.

1-4th Scenario. Driver B is informed that one child and pregnant woman (no.1) from his right side moving from back to forward using S4 and S7 and one old man (no.2) from his right side moving from forward to back using S3 sounds are crossing the intersection. The three pedestrians are represented from right side at the same time to check if driver has enough ability to distinguish three different directional sounds at the same time in one side or not.

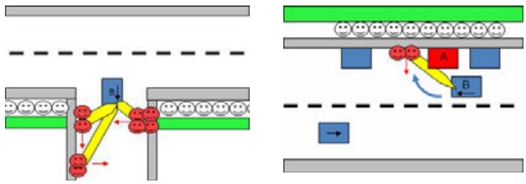


1-3rd Scenario 1-4th Scenario

Fig. 3.

1-5th Scenario. Driver B is informed that so many pedestrians are walking around him using S6 sound. All pedestrians are represented at the same time using all speakers equally to check if driver prefer to have crowd sound from all directions equally or specific directional sound from each side.

1-6th Scenario. Driver B is informed that one child and man are crossing the road from his front side moving from right to left using S1 and S4 sounds at the same time.



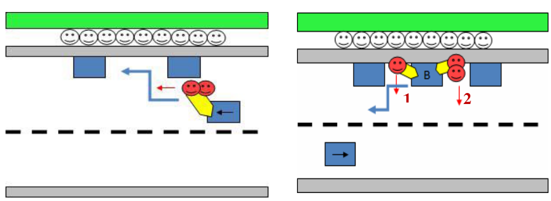
1-5th Scenario 1-6th Scenario

Fig. 4.

Second Set of Scenarios

2-1st Scenario. Driver B is informed that two men are walking along the main road from his right side moving from back to forward using S2 sound. The pedestrians are represented at the same time to check if the driver has enough ability to distinguish the number of the pedestrian or not.

2-2nd Scenario. Driver B is formed that one old man (no.1) from his front side moving from right to left using S3 and one child and pregnant woman (no.2) from his back side moving from right to left using S4 and S7 sounds are crossing the main road while the driver is getting out of park place. All pedestrians are represented at same time.

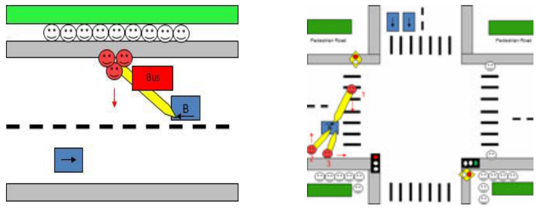


2-1st Scenario 2-2nd Scenario

Fig. 5.

2-3rd Scenario. Driver B is informed that the children are crossing the main road from his front side moving from right to left using S5 sound.

2-4th Scenario. Driver B is informed that one man (no.1) from his front side moving from left to right using S1 and one old man (no.2) from his back side moving from right to left using S3 sound and one man (no.3) from his right side moving from back to forward using S1 sound are crossing the intersection. All pedestrians are represented at same time to measure the ability of driver in distinguishing the moving directions at intersection.

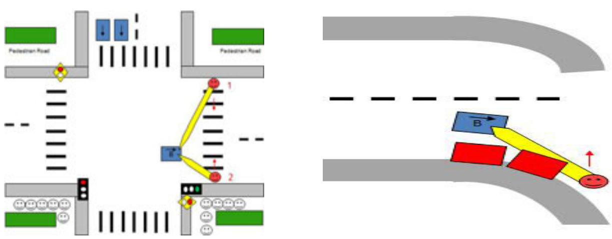


2-3rd Scenario 2-4th Scenario

Fig. 6.

2-5th Scenario. Driver B is informed that two men, (no.1) from his front side moving from left to right and (no.2) from his front side moving from right to left using S1 sound are crossing the intersection. This scenario includes car audio system in which the car audio system is decreased when the 3D natural sound represents the situation. In this scenario the effect of car audio system is tested to check if it can be a cause of a problem in hearing the 3D sound or not.

2-6th Scenario. Driver B is informed that one child is crossing the curve from his front side moving from right to left using S4 sound. This scenario includes car audio system in which the car audio system is decreased automatically when the 3D natural sound represents the situation. In this scenario the effect of car audio system is tested to check if it can be a cause of a problem in hearing the 3D sound or not.



2-5th Scenario 2-6th Scenario

Fig. 7.

3 Result and Discussion

3.1 First Part: The Representative of the Sounds

Figure 8 from first test shows the ability correct perception of the sound that representing the type of pedestrians. Among 15 subjects, 89.875% of users gave right answers and 10.125% of them gave wrong answers. The results showed the design of the natural sounds have good representative to the objects that it represents.

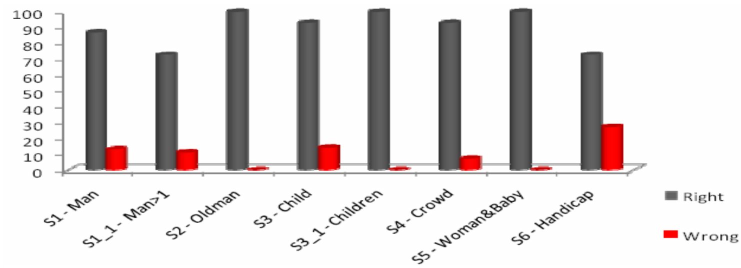


Fig. 8. First part - the representative of the sounds

3.2 Second Part: The Direction of the Sounds

In this part, the perception of the sound's moving direction were measured. Fig 8 shows the data from different directions in which the pedestrians move on. According to this diagram, all subjects were able to distinguish the direction of the sounds DS1, DS2, DS3, DS5, DS6 and DS10 easily.

In the sound test DS4, the man and old man sounds were played at same time but most of subjects were able to distinguish the man sound easier than old man sound because there were some confusion between man and old man steps sound. The old man steps sound were hidden in man steps sound, so the subjects were able to recognize the old man sound just through old man coughing sound, not through old man steps sound and coughing sound together.

Comparing the sound DS10 (the crowd sound without movement) with the sound S8 (the sounds which were played from all the directions) was one of the aims of this test. The data shows that the most of subjects prefer to have crowd sound from all directions equally instead of having specific directional sound for each side due to presenting the crowd situation. In this level, 75.5% of subjects gave correct answers and 24.5% of them went for wrong answers, this means that the directional sounds were enough qualified to draw the drivers' attention to moving direction of pedestrians and the type of them.

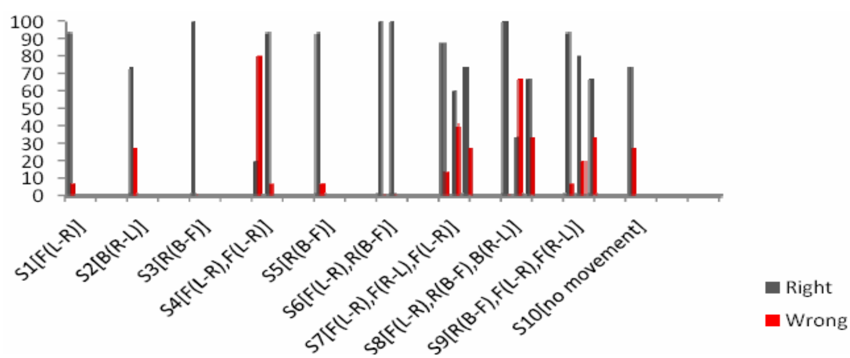


Fig. 9. Second part - the direction of the sounds

3.3 Third Part: The Application of 3D Sounds

First Set of Scenarios

In this part, three questions were asked according to users' driving performance in simulator. First question was about helping the driver perceiving the moving direction of pedestrians. The sounds helped 65% of subjects in 1-3, 1-4, 1-5 and 1-6 scenarios. In the 1-1 scenario, the sound could not help with direction perception and it acted like alarm since the driver can see the pedestrian in front of them. In the 1-2 scenario, the most of subjects have some kind of confusion in distinguishing the direction of all sounds and they preferred to have a crowd sound compare to fifth scenario. In this part, 63.33% of the users gave the answer yes, 34.33% of the users gave the answer no and 2.33% of users selected the answer maybe. The second question was about allocation of attention, the sounds drew the attention of the most subjects in all scenarios. As the results, 84.16% of the users gave the answer yes, 14.66% of the users gave the answer no and 1.16% of users selected the answer maybe. The third question was about the car's engine sound in which 3D sounds can be heard significantly from background engine sound and according to the results engine sound could not be a cause of a problem for designer 3D sound system. 3D sounds and engine sound has same format, frequency and bit ratio but 3D sound was louder than engine sound. In this part, 97.66% of the users gave the answer yes, 2.33% of the users gave the answer no and 0% of users selected the answer maybe for all scenarios.

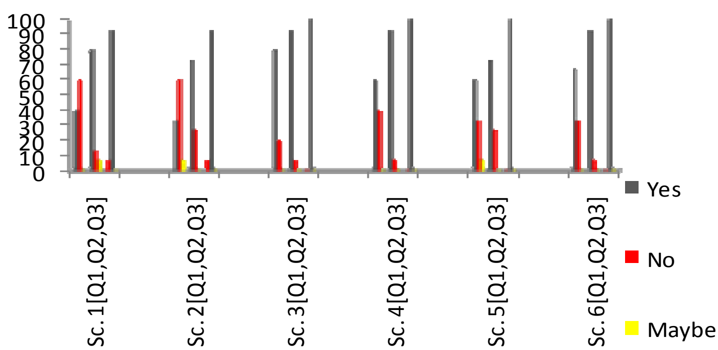


Fig. 10. Third part - First Set of Scenarios - the application of 3D sounds

Second Set of Scenarios

In the figure 11, the first question was about pedestrians moving direction perception and the sounds helped the most of subjects in all scenarios. 78.66% of the users gave the answer yes, 19% of the users gave the answer no and 2.33% of users selected the answer maybe. The second question was about attention allocation. The results showed that sounds drew the attention of the most users in all scenarios, 92.16% of the users gave the answer yes, 7.83% of the users gave the answer no and 0% of users selected the answer maybe. The third question was about the car's engine sound in which 3D sounds can be heard significantly from background engine sound and according to the results, engine sound could not be a cause of a problem for designer 3D sound system. In this part 100% of the users gave the answer yes, 0% of the users gave the answer no and 0% of users selected the answer maybe. The fourth question was about the car audio system's sound in which 3D sounds can be heard significantly from background audio sound and according to the results, audio system sound could not be a cause of a problem for designer 3D sound system. In this part, 83% of the users gave the answer yes, 17% of the users gave the answer no and 0% of users selected the answer maybe.

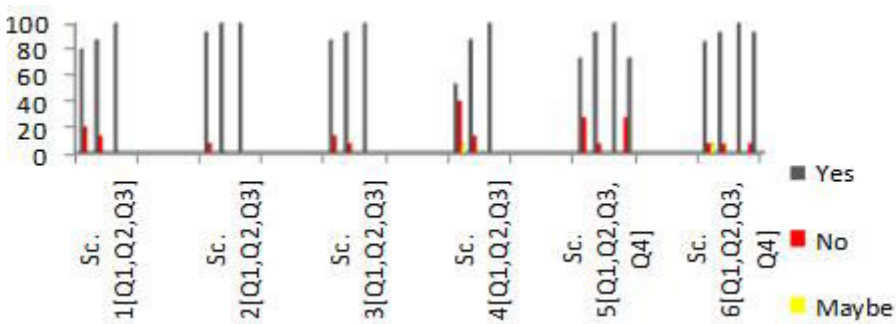


Fig. 11. Third part - Second Set of Scenarios - the application of 3D sounds

4 General Discussion and Future Work

The main reasons for accidents involving car and pedestrian are driver's inattention created by various factors on the road such as crowds, parked vehicles and different information and communication devices that are used in car such as navigation system, audio systems. In order to decrease visual loading for drivers and increase the traffic situation awareness, we study on using different type of natural sound to present different kinds of pedestrian on the road to lead the driver's attention to the right direction and objects. Three series studies are reported in this paper include traffic problems' analysis, the design of natural sounds and simulator evaluation. We design the 3D audio system based on human capability to detect and determine the position of sound sources and according to these studies, using six types of 3D natural sounds all based on most common property (human steps sound) due to reduce the number of sound to present traffic information to driver has strong and positive result instead of having a lot of sounds and it can reduce the driver's confusion, but still we have some unsolved research questions which a lot of more researches need to be carried out on them to implementing the technology into the car in order to warn the driver in good timing.

"Is it necessary to present the traffic information of driver's back while the driver pushes forward and vice versa'' is the first unsolved research question. Reducing the number of 3D awareness sounds also can be a great solution in order to increase the driver's ability to detect the sound sources.

The second unsolved research question which can be consider as a non critical problem is the driver's head movement in the time of playing the natural sounds that can make small distraction in detecting the exact position of the sound source. The human natural sounds were designed according to the position of driver's seat.

In conclusion, using 3D natural sounds to present the pedestrian position to the driver can be very useful system to increase the drivers' awareness of traffic situation. Collected results indicate that the natural sounds can draw the attention of drivers to the specific pedestrian and it guides the attention of drivers towards right direction.

References

[1] Gandhi, T., Trivedi, M.M.: Pedestrian Protection Systems: Issues,Survey, and Challenges. IEEE Transactions on Intelligent Transportation Systems 8(3) (September 2007)

[2] Chen, F., Hargernäss, J., Lindh, M.: Listening to the Traffic – Introduce 3D-Sounds into Truck Cockpit for Traffic Awareness (2009)

[3] Chen, F., Qvint, G., Jarlengrip, J.: There Are Other Road Users Close to You – Improving traffic awareness of truck drivers. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 323–329. Springer, Heidelberg (2007)

[4] Dell, W.: The use of 3D Audio to Improve Auditory Cues in Aircraft, University of Glasgow, Department of computing Science, Glasgow, UK, Class: CS4H/Session: 1999/2000

[5] Haas, E.C.: Can 3-D auditory warning enhance helicopter cockpit safety? In: Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, Illinois, October 5-9 (1998)

[6] Tan, A.K., Lerner, N.D.: Acoustic Localization of the In-Vehicle Crash Avoidance Warning as a Cue to Hazard Direction, Department of Commerce NTIS (1996)

[7] Withington, D.: The use of directional sound to improve the safety of auditory warning. In: The XIVth Triennial Congress of the International Ergonomics Association & 44th Annual Meeting of the Human Factors and Ergonomics Society, San Diego, July 29-August 4 (2000)

Designing Pervasive Games for Learning

Carmelo Ardito[1], Rosa Lanzilotti[1], Dimitrios Raptis[2], Christos Sintoris[2], Nikoleta Yiannoutsou[2], Nikolaos Avouris[2], and Maria Francesca Costabile[1]

[1] Dipartimento di Informatica, Università degli Studi di Bari, Via Orabona 4, 70125 Bari, Italy
[2] HCI Group, University of Patras, Ypatias Str., GR-26500, Rio Patras, Greece
{ardito,lanzilotti,costabile}@di.uniba.it,
{draptis,sintoris}@ece.upatras.gr,
{avouris,yiannoutsou}@upatras.gr

Abstract. Pervasive games have been proposed as a suitable way to support learning, especially in places rich in information, as for example museums and cultural heritage sites. This paper reports on the work performed to identify guidelines that help designers in developing games able to provide an effective learning experience in such contexts. Such guidelines complement other proposals available in the literature. The presented contribution is a first step of a wider work aimed at deepening our understanding of pervasive educational games, with a special emphasis on games in the cultural heritage domain, in order to inform the designers of such challenging applications.

Keywords: Guidelines, educational pervasive games, design.

1 Introduction

A substantial amount of work has been carried out during the last years on designing applications that support people learning in sites of cultural heritage. Most applications exploit web technology to provide information about museums; historical sites, specific exhibitions, and also the so-called “intangible cultural heritage” (e.g. see [6]). The advent of mobile technology has pushed towards systems to be used by museum visitors. The first systems used portable devices without connections to each other and had no context-aware capabilities. By exploiting the latest technological developments, i.e. the miniaturization of computer devices, their increased processing power and their networking capabilities, latest generation systems for supporting museum visits go beyond electronic guides and become visitors’ multimedia companions, which not only provide useful information, but also aim at improving the overall visitors’ experience [3, 27].

A new kind of activity is represented by *pervasive games*: they expand the notion of game in space and time by exploiting mobile devices like smartphones and PDAs with positioning capabilities (e.g., GPS) and other locative media. Alternative terms used for pervasive games are location-based games, augmented-reality games. These games may be played outdoors or indoors and can be compelling for young players as well as for adults [23]. Four main characteristics contribute to the pervasive games’ appeal and to the players’ emotional involvement: [14]: 1) physical experience; 2)

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 99–108, 2011.

mental challenge; 3) social experience; 4) immersion. Specifically, the physical experience pertains to what it is felt when interacting with real and tangible objects together with virtual elements. Moreover, players enjoy additional mental stimuli by having to solve riddles or to perform tasks. Pervasive games require people to meet, socialize and combine their efforts in order to be more effective while playing, thus providing a wider social experience. Finally, the feeling of immersion in the game setting is the main entertainment factor. From the point of view of mobile learning, which focuses on the enriched interaction with context [25], these four characteristics seem to make pervasive games suitable vehicles of learning activities. Thus, a pervasive game designed to support learning would involve: a) structuring the mental challenge around the physical experience with the tangible and virtual objects in question (e.g. museum exhibits); b) integrating in this interaction the social experience through collaboration or competition with others physically or virtually present; and c) using immersion in the game as a means of engagement and motivation for learning.

Pervasive games have been recently proposed to support visits not only to museums, but also to archaeological sites and historical cities [7, 15, 26]. They often have educational goals, aiming to combine learning with fun. It has been shown that these games are able to stimulate students and engage them in their learning activities by requiring different skills to be deployed simultaneously [7, 15, 26].

The increasing importance of pervasive games in the cultural heritage domain calls for increased support for their designers. This paper contributes to this goal by presenting a set of guidelines that will help designers of pervasive games, with a special attention to games to be played at cultural sites.

Next section provides the motivation for design guidelines for educational pervasive games by referring to related work. Next, section 3 describes the methodology adopted for identifying the guidelines, and reports the final set of guidelines, grouped in five dimensions. The paper ends with some final remarks on the use and future development of the presented framework.

2 Related Work

Literature reports several sets of heuristics that have been identified for designing/evaluating games. These heuristics are often quite disparate, even though, in some cases, they address common issues [17]. Initially, researchers concentrated on heuristics for educational games; then, heuristics for video games were proposed; since 2008, educational games are again becoming an important research issue and new heuristics are being postulated.

One of the first researchers proposing heuristics for games was Malone in the 80s; he identified three basic principles: *challenge*, *fantasy* and *curiosity* [21]. Malone also highlighted the importance to evaluate the game content. Later, Lepper and Malone investigated the most important factors for engaging educational games: *challenge*; *balance* between easy and difficult tasks in order to stimulate learners; *fun* activities that help learners address and revise their misconceptions [20]. More recently, Garris et al. [11] examined the literature on educational games and classified the factors that are important to their learning effectiveness. The framework they defined revealed that the motivation "to play and play again" is a key feature of the best educational games and that *feedback* is very useful in learning. On the other hand, Federoof compiled the first playability heuristics, that are very similar to the heuristics defined by Malone; he determined *gameplay* as the most important part of game design, with storytelling, graphics, and sound as auxiliary factors [10].

Looking at heuristics for designing and/or evaluating video games, Desurvire and her colleagues proposed a set of playability heuristics, called Heuristics for Evaluting Playability (HEP), specifically identified to evaluate video, computer and board games [8]. Such heuristics are useful for addressing problems and challenges related to game play, which is an important component of educational games.

In 2007, Korhonen and Koivisto proposed heuristics for mobile multiplayer games [16]. Their approach in identifying these heuristics is similar to ours, since it is based on the evaluation of three different multiplayer mobile games and on the review of existing literature. Seven heuristic were identified that highlighted the importance of *communication*, *collaboration* among players, the *minimization of deviant behaviour*, the *amount of multi-players* involved in the game, the *visibility* of other players, the *social interactions* in groups and communities, and, finally, the importance of a good *network connection* which is a relevant part in any online mobile game.

Wetzel et al. defined a set of guidelines for designing augmented reality games [29]. They analyzed three different games with the goal to identify what is needed to create good mobile location-sensitive games and what causes them to fail. The twelve guidelines focus on the inclusion of 3D features in such games and very marginally consider more general game design aspects.

Pinelle et al. published game usability heuristics based on usability inspections of 108 different video games [24]. They developed ten usability heuristics and many of them are very similar to the Nielsen's heuristics, e.g. *consistency* and *standards*, *visibility* of system status and *help* and *documentation*. Other heuristics are new but they are specific for video games used for entertainment; as the authors are not interested in educational aspects of games.

In our work, we are very much interested in studies that propose heuristics for designing or evaluating educational games. For example, Barnes et al., on the basis of the results of two exploratory studies performed on their prototypes of Game2Learn, a game that teaches introductory computer concepts, provided some important features to be considered to develop effective educational games [4]. The results of their studies highlighted the importance of providing appropriate feedback, that is particularly important in the case of educational games, and of motivating students to stay engaged enough to learn. They also declared that in-game rewards and punishments are vital to the motivation and potential learning of the students. As it will be outlined in the following sections, in our work we come to similar conclusions.

Bellotti et al. in the 2008 proposed a set of heuristics for educational games that exploit virtual reality [5]. Based on the results of informal tests on prototypes of a game that supports players in discovering/investigating historical/artistic details related to a virtually reconstructed area with which they interact, the authors provided some guidelines. However their results may be extended to any type of educational game, i.e. not specifically virtual reality based games. Indeed, examples of such guidelines: games should not be too long and they should be focused on a specific educational purpose, they should allow players to quit games at any time, game scores should be consistent with their difficulty and educational value. In our study we have also included similar guidelines in the set we defined.

In this paper, we propose guidelines that address a more wide view of mobile educational games and complement existing guidelines, found in the literature.

3 Methodology

In order to define guidelines that can help the design of educational pervasive games, we followed a systematic approach, which is inspired by the "case study methodology", an empirical inquiry that investigates a contemporary phenomenon within its real life context using multiple sources of evidence [30]. Case studies are useful to understand some particular problems or situations in great-depth. The case study methodology enables researchers to gain multi-perspective view of a certain phenomenon or series of events and can provide a thorough picture, since many sources of evidence are used [13]. Result generalization is also possible when findings are replicated in multiple case studies.

The first phase of our research was an extensive review of the literature on the topics of interest, namely pervasive games for exploring cultural heritage sites (museums, archaeological parks, historical cities, etc.) and guidelines for game design and evaluation. Several discussions with colleagues and experts in designing pervasive learning games were carried out. Then, the work concentrated on the analysis of three case studies. As a result, a large set of issues relevant for such games were defined. From these issues, preliminary guidelines were proposed. In a successive phase, such guidelines were refined and reduced in number. They were validated by involving some experts, as described in more detail in the rest of this paper. The resulting 36 guidelines, classified along 5 dimensions, are described in Section 3.2. Finally, designers were asked to create their own games by exploiting the proposed guidelines in order to validate the communicability and effectiveness of the guidelines. In the following, the performed activities are reported in more detail.

3.1 Defining the Guidelines

A team of three researchers, experienced in the design of educational pervasive games but with different backgrounds, was formed. They initially investigated three specific pervasive games for the cultural heritage domain by analysing published papers. Their initial goal was to identify as many issues as possible, related to the design of such games. The researchers' knowledge of the design process allowed them to reflect on their own experience, recount important details, highlight different understandings of the design practice, participate in in-depth discussions, and elaborate on issues and concerns.

Each one of the three researchers was assigned the same set of six papers to analyse independently. The papers reported the design and the evaluation of three pervasive games: Explore! for visitors of an archaeological park [2, 7], MuseumScrabble for museum visitors [26, 28] and Frequency 1550 for visitors of a city centre [1, 15]. All three games have the goal of supporting informal learning and utilise a different range of multimedia features, technologies and interaction techniques, in three different settings. Each researcher identified and reported a long list of game issues and then worked independently to clarify them and to eliminate redundancies. A total of 317 issues were reported by the three researchers.

The researchers met to analyze the initial list of 317 items from which a final list was produced, containing 94 issues related to the design of pervasive educational games. The process included identifying similar issues, merging related or duplicate issues, thus refining the whole list. The overall goal was to identify close issues in the final list, which could be addressed by a same guideline. For instance, some of the 94 final issues were: “Competition as a means to increase motivation”, “Competition by hampering the other teams”, “Competition for limited resources”, “Competition for limited resources to induce collaboration (negotiate roles, discuss strategy)”, “Competition for other team’s points,” “Allow competition”, “Force Competition”. These seven issues are covered by guidelines 3.4 and 5.3 in Table 1.

It is useful to organise a set of guidelines along dimensions, in order to support designers in realising important aspects related to design quickly. The process of defining these dimensions included an individual study phase, where the three researchers organised all 94 issues in subsets addressing a certain dimension, and a consolidation phase where the final set of dimensions was defined by combining the results of the individual work. The resulting five dimensions are:

1. *Game General Design*, which refers to issues related to the overall game design process;
2. *Control/Flexibility*, which is a basic dimension of system usability, that with respect to the games considered in this paper, also refers to helping players to be aware of the effects of their choices on the game execution;
3. *Engagement*, which informs on how to provide an experience that captivates the players, by providing hints on how to structure the game, which tools to adopt, etc.;
4. *Educational Aspects*, which informs on interweaving of learning content into the game context, so that the game can have a valid learning influence on the players;
5. *Social Aspects*, which concerns the interaction among the players, role allocation etc. (the underlying assumption is that social activity, e.g. competition, can act as a motivational factor).

Each of the three researchers was provided with a table containing the 94 game issues, organised according to the five identified dimensions. They first worked individually and defined design guidelines that emerged from the issues. The guiding principle for this activity was the need to identify “a set of guidelines that could guide designers who had the task to build a pervasive game, which aims at improving the learning experience of people while visiting cultural heritage sites”.

Prior to the joint refinement process, each researcher compared the set of guidelines he identified with those of the other researchers. Finally, in a discussion and negotiation phase, they consolidated their guidelines in a single set. As a result, 40 guidelines organised in five dimensions were defined, presented next.

3.2 Validating the Guidelines

The first version of dimensions and guidelines, organized in table similar to Table 1, was submitted to four external HCI researchers with experience in the design of educational pervasive games. The main goal was to check if the formulation of the

Table 1a. The final set of 36 design guidelines organized in 5 dimensions

Dimensions		Guidelines
Game General Design	1.1	Exploit metaphors from real-life games, activities, stories
	1.2	Minimize the changes to the physical places (e.g. modifications to the physical structure, installation of special equipment like projectors, big displays, etc.)
	1.3	Create a multidisciplinary design team (including e.g. HCI, cultural heritage, educational experts)
	1.4	Perform formative evaluations and pilot studies to check if tasks' difficulty is appropriate for the intended players
	1.5	Consider the social conventions of the place (e.g. not laughing in a church)
	1.6	Consider to extend the game experience beyond the game session (e.g. participating in a web community)
	1.7	Consider to include activities/events that are not part of the game, but happen in the real world (e.g. the ceremony of change of the guard at noon)
	1.8	Consider to include a game master (e.g. tutor, supervisor, coordinator) and her role: e.g. enforcing the rules, narrating the story
Control / Flexibility	2.1	Let players become familiar with the equipment and the game rules/structure (e.g. by including an introductory phase)
	2.2	Facilitate game learnability (i.e. tasks, rules, constraints, etc. should be easy to understand and to learn)
	2.3	Player should be free to switch between different tasks
	2.4	Reflect on whether to allow players to correct their mistakes: it could be useful to force them to evaluate the consequences of their actions
	2.5	Provide help or hint mechanisms to assist players
	2.6	Consider to provide increasing difficulty levels (either automatic adaptation or human-generated adaptation)
	2.7	Prevent rule breaking by either discouraging it (e.g. with penalties) or by incorporating cheating into the game
	2.8	Make clear the game goal/s (e.g. earning points, completing tasks, being the winner)
	2.9	Make clear the game ending condition/s (e.g. maximum time, target score, end of resources, ...)
	2.10	Consider to provide alternative ways for performing a task or completing the game
	2.11	Make clear the goal of each task and its effects on the overall game
	2.12	Provide immediate feedback about task execution showing its impact on the overall game

guidelines could be misinterpreted. Based on this feedback, some guidelines were rephrased and the final list of 36 guidelines, reported in Table 1, was produced.

We have performed a further informal study by providing a group of HCI students, engaged in game design projects, with the guidelines, requesting them to design a new mobile game or to evaluate their game design work that they had in progress. They reported that, thanks to the guidelines support, they trusted to have addressed important game design issues. More importantly, some students said that the guidelines helped them to make decisions on key points on which they were in doubt about.

We are now planning a more systematic study in order to involve a wide number of designers in the validation of the proposed guidelines. To this aim a website is being created at the web site of the first author; it reports the motivation of the research, the adopted methodology, the guidelines and the dimensions identified. Registered people, navigating through the pages of the website, can read the comments of other

Table 1b. The final set of 36 design guidelines organized in 5 dimensions

Dimensions		Guidelines
Engagement	3.1	Consider to integrate a back-story that is at the basis of game tasks
	3.2	Consider to exploit role-playing (i.e. impersonating a character) to meaningfully link tasks to the back-story (if any)
	3.3	Provide contextual cues linked to specific places or events to convey additional information (e.g. sounds reproducing noises of daily activities in an ancient city)
	3.4	Consider to allow players to interfere to competitors, e.g. stealing/acquiring points
	3.5	Let players practice different skills by including in the game a variety of tasks, such as: perform a quest, identify/visit certain locations, shoot a picture from a specific angle, videotape a route, search for a certain object, perform a certain action/gesture, search/identify a physical mark, answer a question, collect and classifying material
	3.6	Minimize the interaction with the game tools. Players' attention should be focused on the game and the environment
	3.7	Tune the level of awareness of other players' activities (hide/provide/delay information, e.g. showing the score and the progress of the competitors)
	3.8	Consider to include rewards in order to improve players' motivation/satisfaction (e.g. providing multimedia information as a prize for a successful task); integrate rewards tightly with the game tasks and back-story; consider when to provide the rewards to the players (during/after the game)
Educational Aspects	4.1	Consider to include a pre-game activity to prepare players (e.g. some lessons in classroom explaining the historical context in which the game is set)
	4.2	Game should emphasize either vertical or horizontal exploration of a place/topic, i.e., deeply exploring a limited space (or few objects or a specific topic) *vs.* more superficially exploring a broad space (or many objects or several topics)
	4.3	Tasks should require players to link areas, locations, physical objects to concepts, topics, etc.
	4.4	Balance between competition and knowledge acquisition. Too much competition may have a negative impact on knowledge acquisition
	4.5	Include a debriefing phase after the game to allow players to reflect on the game experience. Design it as an individual/collaborative game/activity that supports players to clarify and consolidate the game experience
Social Aspects	5.1	Team players (if any) should be selected based on players' social relations (e.g. friends to maximize collaboration) or according to their skills. Involve in this process a person that knows them very well (e.g. a teacher)
	5.2	Assign responsibilities and tools (e.g. mobile devices, maps, etc.) among team members to induce collaboration. Consider to force, forbid or allow responsibilities exchange among team members
	5.3	Consider to permit, force or neglect the competition among players/teams

people and insert their own. Each guideline has an explanation and/or a concrete example. By only giving the possibility to provide comments could lead to shallow responses; thus, a set of simple questions is provided to induce people to reflect more deeply: Is this guideline important? Is the phrasing correct/clear/understandable? Is it in the right dimension? Is it wrong? Do you expect that thinking about this guideline will contribute positively to the game design?

Researchers to be involved in this larger validation will be selected in order to have people with practical or theoretical background in mobile design, user experience, games, pervasive games, educational games, serious games, etc. To obtain contact information for such researchers, we used search engines on the Internet and our own knowledge of pervasive games designers.

The next step of the validation study will consists in the analysis of the comments posted in the website. This will help us to collect new elements for further discussions. The analysis of the interactions with the website will also allow us to identify the more active researchers among those that posted comments; such researchers can be later interviewed.

The purpose of the interviews is to gain more insight into the comments that have been expressed. Each selected researcher will undergo a semi-structured interview through a recorded Skype call. In order to do not have an interviewer biased by his/her previous experience, we have chosen an HCI researcher expert in carrying out interviews, but not directly involved in this study.

4 Conclusion

Pervasive games have been recently proposed to support visits to cultural heritage sites, such as museums, archaeological parks, historical cities. These games often have educational goals, i.e. they aim at supporting young students learning about history while having fun. Studies show that these games are indeed able to motivate students and effectively engage them in their learning activities [7, 15, 26].

The guidelines proposed in this paper offer insights on the issues that are relevant when designing educational pervasive games; they were defined by following a systematic methodology. Some guidelines we found in literature are similar to those proposed by us. This fact further validates our study, as the literature has not influenced the process of identifying our guidelines, which were produced in a bottom up approach from identified issues in three typical pervasive games. Our effort has been to integrate and to organize them in a unique set to be more operational for designers of pervasive games.

Even if we are motivated by pervasive games in the cultural heritage domain, the identified guidelines are quite general and may be used for educational pervasive games independently of the specific place of the game and the field of learning.

Current work consists in further validating and refining the proposed guidelines through more systematic studies involving a wider number of designers.

Acknowledgments. The financial support of Italian MIUR through the "L4ALL" grant is acknowledged. Partial support has been provided by the EU COST Action IC0904 Twintide under the “Short Term Scientific Mission” scheme, which funded the short term visit of Carmelo Ardito to the HCI Group of the University of Patras.

References

1. Akkerman, S., Admiraal, W., Huizenga, J.: Storification in History education: A mobile game in and about medieval Amsterdam. Computer Education 52(2), 449–459 (2009)
2. Ardito, C., Buono, P., Costabile, M.F., Lanzilotti, R., Piccinno, A.: Enabling interactive exploration of cultural heritage: an experience of designing systems for mobile devices. Knowledge Technology and Policy 22(1), 79–86 (2009)

3. Barbieri, G., Celentano, A., Orsini, R., Pittarello, F.: Understanding art exhibitions: from audioguides to multimedia companions. In: DMS 2009, pp. 250–255. Knowledge Systems Institute, Skokie (2009)
4. Barnes, T., Powell, E., Chaffin, A., Lipford, H.: Game2Learn: improving the motivation of CS1 students. In: GDCSE 2008, pp. 1–5. ACM Press, New York (2008)
5. Bellotti, F., Berta, R., De Gloria, A., Zappi, V.: Exploring gaming mechanisms to enhance knowledge acquisition in virtual worlds. In: DIMEA 2008, pp. 77–84. ACM, New York (2008)
6. Buono, P., Di Bitonto, P., Di Tria, F., Plantamura, V. L.: Genòmena: a Knowledge-Based System for the Valorization of Intangible Cultural Heritage. In: DMS 2009, pp. 100–105. Knowledge Systems Institute, Skokie, IL, (2009)
7. Costabile, M.F., De Angeli, A., Lanzilotti, R., Ardito, C., Buono, P., Pederson, T.: Explore! Possibilities and challenges of mobile learning. In: CHI 2008, pp. 145–154. ACM, New York (2008)
8. Desurvire, H., Wiberg, C.: Master of the game: assessing approachability in future game design. In: CHI 2008, pp. 3177–3182. ACM, New York (2008)
9. Eriksson, D., Peitz, J., Björk, S.: Socially Adaptable Games. In: DIGRA 2005, DiGRA Digital Library (2005)
10. Federoff, M.: Heuristics and Usability Guidelines for the Creation and Evaluation of FUN in Video Games. Thesis at the University Graduate School of Indiana University (2002)
11. Garris, Ahlers, Driskell: Games, motivation, and learning: a research and practice model. Simulation & Gaming 33(4), 441–467, 2002
12. Ghellal, S., Bullerdiek, S., Lindt, I., Pankoke-Babatz, U., Adams, M., Söderlund, T., Oppermann, L.: Design Guidelines for Crossmedia Game Production. Public IPerG Deliverable D8.1, `http://www.pervasive-gaming.org/Deliverables/D8.1-Design-Guidelines-for-Crossmedia.pdf` (last accessed on January 16 2011)
13. Gummesson, E.: Qualitative Methods in Management Research. Sage Publication, California (1991)
14. Hinske, S., Lampe, M., Magerkurth, C., Röcker, C.: Classifying pervasive games: on pervasive computing and mixed reality. In: Concepts and Technologies for Pervasive Games - A Reader for Pervasive Gaming Research, vol. 1. Shaker Verlag, Aachen (2007)
15. Huizenga, J., Admiraal, W., Akkerman, S., ten Dam, G.: Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game. Journal of Computer Assisted Learning 25(4), 332–344 (2009)
16. Korhonen, H., Koivisto, E.M.I.: Playability heuristics for mobile multi-player games. In: DIMEA 2007, pp. 28–35. ACM, New York (2007)
17. Korhonen, H., Paavilainen, J., Saarenpää, H.: Expert review method in game evaluations: comparison of two playability heuristic sets. In: 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era (MindTrek 2009), pp. 74–81. ACM, New York (2009)
18. Korhonen, H., Saarenpää, H., Paavilainen, J.: Pervasive Mobile Games — A New Mindset for Players and Developers. In: Markopoulos, P., de Ruyter, B., IJsselsteijn, W.A., Rowland, D. (eds.) Fun and Games 2008. LNCS, vol. 5294, pp. 21–32. Springer, Heidelberg (2008)
19. Lazar, J.: Research methods in human-computer interaction. Wiley, Chichester (2010)
20. Lepper, M.R., Malone, T.W.: Intrinsic motivation and instructional effectiveness in computer-based education. In: Snow, R.E., Farr, M.J. (eds.) Aptitude, Learning, and Instruction, vol. 3, pp. 255–286 (1987)

21. Malone, T.W.: Heuristics for designing enjoyable user interfaces: Lessons from computer games. In: CHI 1982, pp. 63–68. ACM, New York (1982)
22. Mitchell, J.C.: Case study and situational analysis. Sociological Review 31(2), 187–221 (1983)
23. Montola, M., Stenros, J., Waern, A.: Pervasive Games: Theory and Design. Morgan Kaufmann, San Francisco (2009)
24. Pinelle, D., Wong, N., Stach, T.: Heuristic evaluation for games: usability principles for video game design. In: CHI 2008, pp. 1453–1462. ACM, New York (2008)
25. Sharples, M., Milrad, M., Arnedillo Sánchez, I., Vavoula, G.: Mobile Learning: Small devices, Big Issues. In: Balacheff, N., Ludvigsen, S., de Jong, T., Lazonder, A., Barnes, S. (eds.) Technology Enhanced Learning: Principles and Products, pp. 233–249. Springer, Heidelberg (2009)
26. Sintoris, C., Stoica, A., Papadimitriou, I., Yiannoutsou, N., Vassilis, K., Avouris, N.: MuseumScrabble: Design of a mobile game for children's interaction with a digitally augmented cultural space. International Journal of Mobile Human Computer Interaction 2(2) (2010)
27. Stock, O., Zancanaro, M.: PEACH - Intelligent Interfaces for Museum Visits (Cognitive Technologies). Springer, Heidelberg (2007)
28. Stoica, A., Fiotakis, G., Raptis, D., Papadimitriou, I., Komis, V., Avouris, N.: Field evaluation of collaborative mobile applications. In: Handbook of Research on User Interface Design and Evaluation for Mobile Technology. Idea Group, USA (2007)
29. Wetzel, R., McCall, R., Braun, A.-K., Broll, W.: Guidelines for designing augmented reality games. In: Future Play 2008, pp. 173–180. ACM, New York (2008)
30. Yin, R.: Case study research: Design and methods, 4th edn. Sage Publishing, Beverly Hills (2009)

Customized Usability Engineering for a Solar Control Unit: Adapting Traditional Methods to Domain and Project Constraints

Patricia Böhm, Tim Schneidermeier, and Christian Wolff

University of Regensburg, Department of Media Informatics, Universitätsstrasse 31, 93053 Regensburg, Germany
Patricia.Boehm@stud.uni-regensburg.de, {Tim.Schneidermeier,Christian.Wolff}@sprachlit.uni-regensburg.de

Abstract. This paper describes the adaption and customization of usability engineering methods for the interface design of a solar control unit. The design of a nontraditional interface, constrained access to representative users and a lack of common interface standards were domain-related issues to overcome. Due to limited resources, a Guerilla HCI approach was established. Traditional low-cost methods like prototyping and simplified usability testing were applied and adapted to fit in the domain-specific context. Good feedback indicates suitability of modified discount methods in the new domain.

Keywords: usability engineering, discount usability, user-centered design, user interface design, nontraditional user interfaces, facility management.

1 Introduction

The advantages of usability testing for the development of well-designed, intuitive and easy to use interactive software have been acknowledged for a long time. Since the rise of smart phones and home entertainment systems, usability engineering found its way into additional application areas [1], [2], [3], [4]. Nevertheless, in many technical domains like facility management usability engineering methods still need to be adapted to fit in the specific context. However domain-specific usability has not been established by research approaches yet [5].

This paper describes a case study of an user-centered design (UCD) approach for designing the user interface of a solar control unit. It focuses on the adaption and customization of established low-cost usability methods to the domain of nontraditional interfaces, i.e. domestic information appliances encompassing its critical constraints and challenges and their influence on the method selection, a more detailed description of the design process and results can be found in [6]. Usability efforts in the field of facility management has to overcome domain-related issues like the design of nontraditional interfaces [7], the lack of common interface standards and limited access to representative users. Constraints of human and financial resources, a common challenge in UCD practice [8], have to be managed. In the context of an interdisciplinary research funded by the Bavarian Ministry of Economic

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 109–117, 2011.

Development[1], the project team consisted of usability engineers of the University of Regensburg and two domain experts of the enterprise emz Hanauer.[2]

In the first section the general approach of customized low-cost usability engineering, it`s goals and constraints in the context of a solar control device are pointed out. In section 3 an approach towards prototyping and evaluating a nontraditional interface is shown. This is followed by a description of a card sorting study dealing with the different capabilities of users especially domain knowledge. The last section describes the approach of parallel usability testing. A modified usability testing procedure, including different tasks and simultaneous testing sessions was established in order to compensate limited availability of users.

2 Customizing Usability Methods

Usability Engineering can be defined as a set of methodologies and activities used for designing an interactive product [8]. Ideally, the most appropriate methods are selected for best results. Yet reality shows that in most projects many parameters like project resources, access to users, interface properties or return of investment (ROI) have to be respected. Approaches towards discount or guerrilla usability are facing this challenge [9], [10], [11], [12]. The main idea is to use simpler and therefore cheaper methods in constrained contexts. For example Nielsen's *discount usability engineering* is based on techniques like scenarios and simplified thinking aloud [13]. Due to limited human and financial resources these ideas were also taken account of in the design of the smart sol control unit. The methods were customized and adapted to the domain of nontraditional interfaces. Prototyping and simplified user testing including the thinking aloud method built the basis for our Guerilla HCI approach.

2.1 Usability Goals and Requirements

User characteristics and typical tasks were identified by informal interviews with domain experts and potential users of the control unit. Main user groups could be defined as installers and end-users. Differing in terms of usage patterns, expertise in control devices as well as knowledge about solar systems in general, both needs had to be addressed. The craftsman on the one hand is responsible for setting up and attendance of the device, being familiar with this kind of devices in as well as with the according terminology. On the other hand, typical end-users hardly use these devices at all due to hard to use and counter-intuitive interfaces of state of the art control devices.[3] They are usually not familiar with technical terms or the general functioning of the solar plant or other devices of comparable technical complexity. According to these requirements, two goals had to be accomplished: On the one hand domain specific workflows have to be optimized, and on the other hand, an intuitive and self-explanatory user interface has to be designed to break down barriers for the end-user.

---

[1] Grant IUK 0910-0003.

[2] http://www.emz-hanauer.de, a medium-sized enterprise specialized in household appliances as well as domestic and environmental engineering, located in Nabburg, Bavaria [accessed February 2011].

[3] Results of market research and user interviews conducted by the authors in 2009/2010.

2.2 Testing Constraints

Besides resource limitations, other constraints influenced method selection and adaption. As described above, usability testing had to involve two user groups: end users and installers. Whereas possible end users are more or less easy to reach, e.g. using student volunteers, recruiting installers as test persons is more challenging. Thus an efficient way of *parallel usability* testing was established to compensate their limited availability.

User interface guidelines and sets of predefined input controls (e.g. mouse, keyboard) limit the design space for desktop interfaces. While each application provides different functionality, consistent controls for standard operations such as opening, closing or setting preferences can be expected. Designing a nontraditional interface involves far more design decisions and challenges including not just the on-screen interface but hardware parts as well. The predictability of consistent controls is completely absent in nontraditional interfaces [14]. Thus we had to modify and extend typical prototyping methods.

3 Prototyping a Nontraditional Interface

The primary method chosen for evaluating the design concepts was prototyping. Because of the fast and comparatively easy process, low-fidelity paper prototypes were applied at first [13], [10]. It turned out that they were fine to represent a quick idea of the on-screen layout, but insufficient in all other means regarding nontraditional interfaces: Hardware interaction, physical properties of controls and visual aspects needed to be sketched.

3.1 Hardware Dummy to Evaluate Overall Design

Due to time constraints especially regarding the production period the device hardware had to be designed first. Based on a competitor analysis and expert reviews on usability and required human-machine interaction, the hardware design was developed in cooperation with an industrial design company. A full color display was

Fig. 1. Overall design concept

chosen to overcome the technical impression generated by reduced display techniques [15]. The interaction is based on a click wheel for navigation and confirmation and an escape button for going back. Click wheels are appropriate for the fast adjustment of parameters with great values ranges [16]. The escape button guarantees the easy reversal of actions as an important principle for interface design [17]. To get an impression of the physical properties an (almost non-functional) hardware dummy was built. In cooperation with domain experts the basic shape and positioning of the control elements were iteratively evaluated and improved. Thus, design flaws like a lack of height of the click-wheel could be revised in an early stage.

3.2 Evaluation of Small-Display Layouts with Mockups

Based on the earlier paper sketches, graphic mockups were used for design and evaluation of the screen layout. In contrast to traditional website or software application wireframes, the layout also had to consider the constraints of small-display characteristics, e.g. visual angle, viewing distance, luminance, character and font attributes, display placement [18]. Thus graphic mock ups were used to communicate our ideas to domain specialists and were reusable for usability testing. Because of the recycling time could be saved. For the evaluation of font, font-size and color schemes, mockups were rendered to imply real size and resolution constraints of the final display.

3.3 Evaluating Interaction with an HTML-Based Prototype

The evaluation of the overall interaction design concept was more demanding. Due to time and resource constraints, the concept had to be evaluated without an operational hardware prototype. In addition, interaction had to be modeled without the final information architecture. Final menu structure determination had to be deferred to a later point as representative users were not available at the time.

Vertical interactive html-prototypes proved to be a valid approach to these problems. Two important use cases, the initial setting up of the device and an error handling situation were implemented to address both user groups, installers and end users. In order to simulate a model as realistic as possible despite the given constraints a virtual representation of the device`s body and controls were included. Hence users could interact with the prototype using a virtual click wheel and escape button.

Fig. 2. Interactive html-based prototype

4 Card Sorting and Domain Constraints

41 menu items were extracted out of the functional specification document of the smart sol controller. Organizing them in a logical and intuitive structure that matches both user groups is crucial for the intended self-explanatory and efficient interface. Therefore, a card sorting method was chosen, a distinguished method to create information architecture according to user's metal model [19]. Originally using real paper cards, the method has been translated into an online exercise with virtual cards [20].

4.1 Failure of Online-Based Open Card Sorting

Due to the domain-specific terminology and the lack of standards an *open card sort* seemed to be appropriate: Here, participants have to group items and label the groups themselves [20]. A first attempt using a free online tool[4] failed. The participants, installers to whom we got in touch through the enterprise as well as randomly chosen students received an e-mail with an invitation and a link to the online tool. The return of installers was very low. Little computer experience and a lack of understanding of the goal might have led to the poor results, probably typical for this kind of design task and test users. In addition, grouping and labeling by end user was very heterogeneous and inconsistent. Therefore, we concluded that finding specific categories for the technical and abstract items is too difficult for non-professionals.

4.2 Customizing Card Sorting to Domain Constraints

In a second try we strived to take the specifics of the different target groups into account. To simplify the task, installers and end-users should participate in a closed card sort where categories are predefined. To generate the hyperonyms for the categories on the highest hierarchy level an open card sort with five domain experts of the enterprise using the online tool was conducted. Their labeling was not unambiguous, but the groupings implied a helpful domain-related view. Additionally, we compared their classifications to the menu structures of more or less well-designed competitor devices. For nomenclature of categories and items we tried to strike a balance between popularity, distinctness and domain conventions. Subsequently student volunteers were kindly asked to take part in the closed card sort using the online tool once more. In order to motivate the craftsmen and take their potential lack of computer affinity into account we established a real paper card sort with the presence of test supervisors who could be asked if problems would occur. In order to save time and make the most out of their voluntary participation, the card sorting session of the installers was combined with other usability tests on one evening. The procedure and its constraints are described in the section *parallel usability testing* below.

4.3 Simple and Quick Analyzing

Analyzing the gained data from the conducted card sort was done right after the session. Data from each group was first evaluated separately before the results were

---

[4] WebCAT: http://zing.ncsl.nist.gov/WebTools/WebCAT/overview.html [accessed February 2011].

assembled in a second step. Basically, three different methods for analysis are available: Analysis with a special card sorting software, cluster analysis with a statistical software or frequency detection with spreadsheets [19].

Although the applied online-tool offers rudimentary analysis functions the results of the paper card sorting had also to be included. Therefor we used SPSS for cluster analysis as well as frequency distribution. According to our experience using spreadsheets is a practicable and efficient way. The main idea is calculating for each card and category the percentages of participants that assigned them. Nevertheless frequencies are insufficient when there is no majority and allocation is ambiguous. Therefore cluster analysis can be used to get supplementary clues. From cluster analysis a dendogram can be derived: a tree diagram where the cards that were sorted the most similar by the participants is placed on branches that are close together. Considering the point at which the items join together their similarity can be deduced [20]. Altogether, spreadsheet approach and cluster analysis complement each other fairly well.

5 Parallel Usability Testing

All usability tests with installers were conducted on one day, as the craftsmen were not jointly available for a longer period of time. The *parallel usability testing* including card sorting, user tests as well as a post-test questionnaire required adapted procedures in terms of recruitment of participants, test environment and material.

5.1 Recruitment of Installers

Usually test persons for a usability study are recruited by special agencies or through advertisements and get some kind of reward for their participation [21], [22]. Having restricted resources we had to search for others ways of recruitment.

Useful contacts to installers could be established using the enterprise`s network and asking them for voluntary participation. To raise willingness for voluntary participation, usability testing was embedded in an informal social event with all participants getting an invitation via mail. The event began with a (rare) guided tour through the production department of the enterprise before usability sessions were started. A get-together with snacks at the end underlined the casual character. Despite these promotional efforts only 11 out of 20 invited installers participated.

5.2 Test Environment for Simultaneous Testing

Though testing in the field is the best method to consider factors of context of use, and therefore detect more severe issues [23], [24], it was not practicable in our approach. Using a HTML-prototype as objective, testing in a traditional laboratory environment was the only option.

Due to their voluntary participation usability testing had to be accomplished in a reasonable time. The testing sessions were conducted simultaneously instead of one after another. Therefore two laboratory-like testing environments were established in office accommodations of the enterprise. One room was prepared with laptops for user tests and another room was equipped with tables and paper card sets for card sorting. For the simultaneous proceeding additional supervisors were needed. Domain experts of the enterprise and five additional employees, mainly engineers and programmers, were briefed and deployed as additional test supervisors.

5.3 Test Tasks and Procedure

In order to make the most out of the voluntary participation of the target group testing sessions included two tasks: Closed card sorting (using real paper cards) with predefined categories and conducting two use cases applying the interactive html prototype. The participants were divided into two groups, each of which accomplished the different tasks in reverse order to minimize statistical impact. To guarantee traceability and to speed up the test process, protocols and documents were prepared in advance for the testing session. Each test person received a portfolio containing instructions, questionnaires and protocol forms already labeled with ids. The card sorting was conducted with two supervisors. After a short introduction the installers were free to start with the sorting. The groupings were left on the table and recorded with a protocol form.

During user tests every participant was supported by a single supervisor and asked to think aloud. All comments were written down in a predefined protocol form. A post-test questionnaire including questions about interaction concept, layout aspects and workflow of test scenarios, completed the test. Altogether parallel usability testing worked out quite well: Engaging domain specialists in usability testing raised efficiency and helped bridging the knowledge gap. Due to their expertise questions on technical details of supported solar systems could be answered. In addition, their domain-related view on actual usability problems was helpful for assessing these issues. No major usability problems were found and the overall interaction design has been approved. The final design of the *smart sol* control unit was rated as intuitive and well-arranged.

5.4 Method Limitations

Due to time pressure user testing was conducted with HTML-based prototypes. Though simulating hardware interaction with virtual elements worked out, this kind of testing objective is limited. Testing the on-screen user interface on the real hardware would for sure have been an enrichment. Nevertheless due to all constraints mentioned above the *quick and dirty* solution turned out fine in this case. Other important factors like the context of use of the device, usually wall-mounted in a basement-like area, could neither be considered in the test sessions.

6 Conclusion

In this paper a customized Guerilla HCI approach used for the development of a solar control with constrained resources is described. Traditional low-cost methods were applied and adapted to fit in the domain-specific context. The *smart sol* solar control was presented at *Intersolar Europe 2010,* the world`s largest trade show for solar products[5]. The overall design as well as the interaction design in particular were rated a success. This feedback and the positive resonance of user testing indicate that discount usability engineering approaches are suited in domains beyond the desktop. Nevertheless our modifications show that there is a need for *domain-specific usability*, an area research is just beginning to explore, e. g. in the *1st European Workshop on HCI Design and Evaluation*[6] focusing on the influence of domains Further studies examining the usability of the controller in the field are planned to enlarge our findings.

Acknowledgments. This work has been conducted in the context of the research project Modino that is funded by the Bavarian Ministry of Economic Development (Grant IUK 0910-0003).

References

1. Hallnäs, L., Redström, J.: From Use to Presence: On Expressing and Aesthetics of Everyday Computational Things. ACM Trans. Comput.-Hum. Interact. 9, 106–124 (2002)
2. Han, S.H., Yun, M.H., Kwahk, J., Hong, S.W.: Usability of Consumer Electronic Products. International Journal of Industrial Ergonomics 28, 143–151 (2001)
3. Petersen, M.G., Madsen, K.H., Kjær, A.: The Usability of Everyday Technology: Emerging and Fading Opportunities. ACM Trans. Comput.-Hum. Interact. 9, 74–105 (2002)
4. Weiss, S.: Handheld Usability. Wiley, New York (2002)
5. Chilana, P.K., Wobbrock, J.O., Ko, A.J.: Understanding usability practices in complex domains. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 2337–2346. ACM, Atlanta (2010)
6. Smart Sol – Bringing User Experience to Facility Management: Designing the User Interaction of a Solar Control Unit. To appear in: Proceedings of HCI International 2011 (2011)
7. Kortum, P. (hrsg): HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces. Morgan Kaufmann, San Francisco (2008)
8. Vredenburg, K., Mao, J., Smith, P.W., Carey, T.: A survey of user-centered design practice. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves, pp. 471–478. ACM, Minneapolis (2002)
9. Kane, D.: Finding a Place for Discount Usability Engineering in Agile Development: Throwing Down the Gauntlet. In: Proceedings of the Conference on Agile Development, p. 40. IEEE Computer Society, Los Alamitos (2003)
10. Snyder, C.: Paper prototyping: the fast and easy way to design and refine user interfaces. Morgan Kaufmann, Amsterdam (2003)
11. Bias, R.G., Mayhew, D.J. (eds.): Cost-Justifying Usability. Academic Press, Inc., Orlando (1994)
12. Nielsen, J., Mack, R.L.: Usability inspection methods. Wiley, New York (1994)

[5] http://www.intersolar.de [accessed February 2011].

[6] http://sites.google.com/site/ehcide/, taking place in Limassol, Cyprus in April 2011 [accessed February 2011].

13. Nielsen, J.: Guerilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. In: Bias, R.G., Mayhew, D.J. (eds.) Cost-Justifying Usability, pp. 245–270. Morgan Kaufmann Publishers, San Francisco (1994)
14. Kortum, P.: Introduction to the Human Factors of Nontraditional Interfaces. In: Kortum, P. (ed.) HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces, pp. 1–24. Morgan Kaufmann, San Francisco (2008)
15. Schmidt, A., Terrenghi, L.: Methods and Guidelines for the Design and Development of Domestic Ubiquitous Computing Applications. In: Proceedings of the Fifth IEEE International Conference on Pervasive Computing and Communications, pp. 97–107. IEEE Computer Society, Los Alamitos (2007)
16. Baumann, K., Lanz, H.: Controls. In: Baumann, K., Lanz, H. (eds.) User Interface of Electronic Appliances, pp. 131–161. Taylor & Francis, New York (2001)
17. Shneiderman, B., Plaisant, C.: Designing the user interface: Strategies for effective human-computer interaction. Addison-Wesley/Pearson, Upper Saddle River, NJ (2010)
18. Kärkkäinen, L., Laarni, J.: Designing for small display screens. In: Proceedings of the Second Nordic Conference on Human-Computer Interaction, pp. 227–230. ACM, New York (2002)
19. Courage, C., Baxter, K.: Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques. Morgan Kaufmann Publishers, San Francisco (2004)
20. Tullis, T.: Measuring the User Experience. Elsevier, Amsterdam (2008)
21. Dumas, J.S., Redish, J.C.: A Practical Guide to Usability Testing. Intellect, Exeter (1999)
22. Rubin, J., Chisnell, D.: Handbook of Usability Testing. How to Plan Design and Conduct Effective Tests. Wiley, Indianapolis (2008)
23. Duh, H.B., Tan, G.C.B., Chen, V.H.: Usability evaluation for mobile device: a comparison of laboratory and field tests. In: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 181–186. ACM, Helsinki (2006)
24. Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., Stenild, S.: It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 272–280. ACM, Oslo (2006)

End-User Composition Interfaces for Smart Environments: A Preliminary Study of Usability Factors

Yngve Dahl[1] and Reidar-Martin Svendsen[2]

[1]SINTEF ICT,
7465 Trondheim, Norway
[2] Telenor Corporate Development,
7004 Trondheim, Norway
yngve.dahl@sintef.no, Reidar-Martin.Svendsen@telenor.com

Abstract. This paper describes a preliminary study of factors that influence the usability of end-user composition interfaces for smart environments. Three early GUI prototypes were tested in a usability laboratory, and transcriptions from the test subjects' comments during the experiment were analyzed in search of recurring areas of concern. Four usability factors were identified: (1) predictability of composition model, (2) readability of composition representation, (3) overview and means for planning compositions, and (4) attractiveness and desirability.

Keywords: End-user composition, Graphical user interfaces, Interface Metaphor, Smart environments, Usability.

1 Introduction

The realization of smart environments is often framed as a technical problem related to the lack of mechanisms that allow heterogeneous devices to exchange data and respond to local events [1]. Realizing smart environments, however, also requires intelligible user interfaces that allow people to customize and control their local surroundings. This paper describes a preliminary comparative study of graphical user interfaces (GUIs) designed to help end-users combine networked resources inside a smart environment into compositions that can offer new extended functionality. The focus is on GUIs that are applicable for design-time composition, that is, specification of compositions in advance of using them. The evaluated GUIs primarily target composition editors running on desktop computers.

The main objective of this paper is to identify factors influencing a composition interface's effectiveness, efficiency, and satisfaction in use. Effectiveness, efficiency, and user satisfaction are key elements in the ISO definition of usability [2]. We refer to factors affecting these three elements as *usability factors*.

Three composition interfaces were prototyped and tested as part of a problem solving exercise. Test subjects were given tasks related to composition of devices and services provided in hypothetical domestic smart environment. The experiment was conducted in a usability laboratory with sixteen test participants (architect and IT students). Transcriptions from the users' comments during testing were analyzed to

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 118–127, 2011.

identify factors that influenced the usability of the composition interfaces, as perceived b the test subjects.

The paper's main contribution is a qualitative understanding of how the usability of end-user composition interfaces is intimately dependent a set of critical factors.

2 Related Work

Alternative approaches to end-user composition have been explored in earlier work. Humble et al. [3] describe an editor that allows users to configure a domestic ubiquitous computing environment. The GUI of the editor is based on a jigsaw metaphor. Users can combine functionalities offered by various components, by coupling graphical puzzle-like pieces that represent the components. Newman [4] proposes a concept where "cooking recipes" form the underlying metaphor. The digital recipes describe the required ingredients (hardware and software components) and preparations (steps) to make a composition. The Speakeasy system [10] employs a web browser interface and task-oriented templates. The Device Composition Aquarium [5] displays compositions as graphs or "wires".

The studies cited above illustrate some of the variety of design metaphors and concepts that have been investigated in the context of end-user composition. Although various approaches have been proposed, there is relatively little practical guidance as to what interface characteristics that influence the usability of various approaches.

3 Overall Research Design

The conducted evaluation is formative in nature. As opposed to measuring user performance (e.g., task solving time, number of slips and breakdowns, etc.) for different user interfaces, our goal was to evoke usability-related reflections among test subjects to help inform further design. We have observed architect and IT students interacting with early prototypes to solve composition tasks, and asked them to articulate their experience from use. This has resulted primarily in qualitative data. Given the limited expressive power of the prototypes (see Sect. 4.1), we chose not to investigate possible correlations between programming experience and preference for certain composition interfaces. Instead we focus exclusively on the qualities and drawbacks of the prototypes as articulated by the test subjects.

The evaluation was conducted in a usability laboratory. The test subjects used a conventional PC with keyboard and mouse to test the composition interfaces. Earlier work [6] has identified PCs as the user preferred tool for specifying automation of tasks that can be predetermined and planned.

To give test participants an initial idea of the basic purposes of the prototypes they were given a brief introduction to a hypothetical domestic smart environment. They were asked to imagine that they were living in a resident where various domestic devices and sensors could offer services, respond to local events, and exchange data over a digital network. To help concretize the setting, the test subjects were shown a paper sheet illustrating the different types of devices, or components, they would be dealing with as part of the composition tasks (Fig. 1). The overview also indicated the assumed physical location of the various components.

Fig. 1. The various components that the test participants would be dealing with as part of the composition tasks were illustrated on a paper sheet

4 Prototypes and Experimental Setup

4.1 Simplicity in Use vs. Expressive Power

In order to conduct a comparative evaluation of end-user composition interfaces it is necessary that all candidate solutions allow the same behavior to be specified. This, again, raises the need for a common composition model.

When developing tools for specifying system behavior *simplicity in use* and *expressive power* form two mutually exclusive criteria [3]. Conventional programming languages maximize expressiveness, but raise the threshold for successful use in doing so.

Given our focus on evaluating early concepts, we chose to limit the expressive power of the prototypes to simple condition-response rules. For example, a composition could specify that if an in-house smoke detector is triggered, an SMS notification should be sent to the house owner's mobile phone. In this case, the execution of a service associated with one device (due to some external event), causes another predefined device to respond by executing another service.

For simplicity, we also decided to omit features such as specification of parameters (e.g., mobile phone number, SMS content) related to different devices and services.

4.2 Prototypes

Inspired by earlier concepts (e.g., [3, 5]), three prototypical composition interfaces were build. The three prototypes were *filtered lists, wiring diagram*, and *jigsaw puzzle*. Below, we describe each prototype in further detail.

Filtered Lists Prototype. This prototype allows users to specify a composition through sequential steps. A composition is defined by subsequently specifying the condition (component and event) and the response (component and service) from respective lists (Fig. 2). Filters are applied so that the lists only present alternatives that are relevant for previous selections. For examples, if a user first selects a smoke detector as the triggering device, only events that the smoke detector can respond to are shown in the events list. Relevant response components and services are only shown after the trigger device and event has been selected.

Saved compositions are shown in a dedicated table.

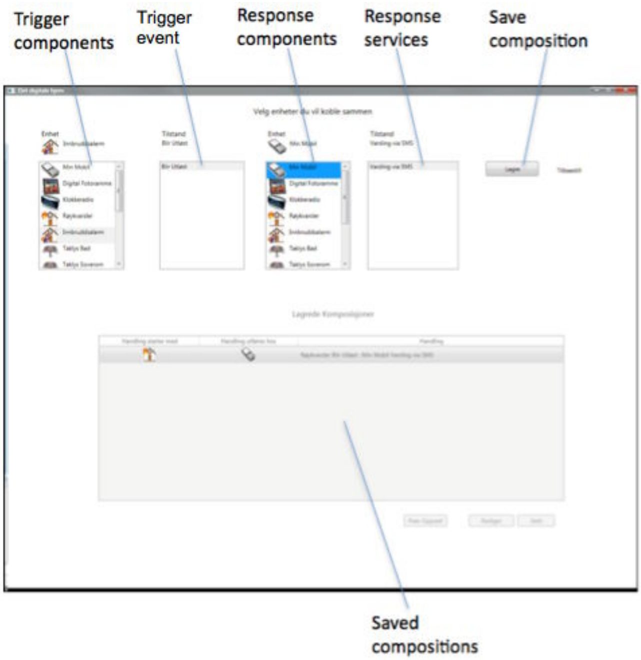


Fig. 2. Filtered lists prototype

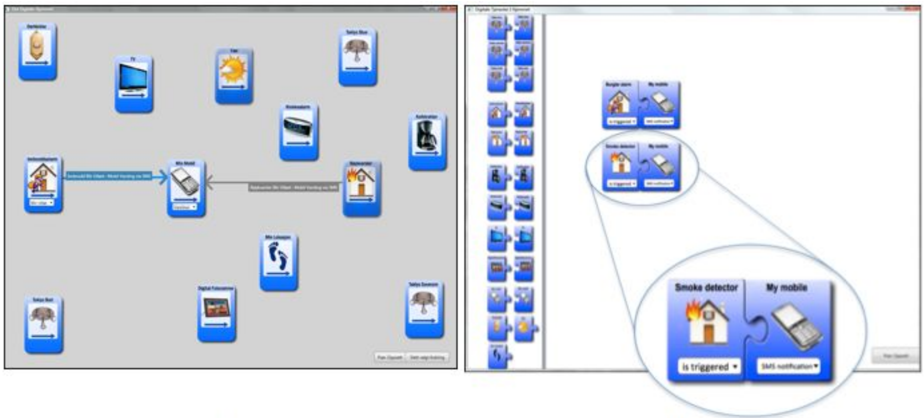
Wiring Diagram Prototype. The metaphor of coupling or "wiring" together separate GUI elements to denote relationships, has been extensively used in different types of web composition tools (e.g., Microsoft Popfly). The metaphor builds on the way stereo system components (tuners, amplifiers and speakers) are hardwired together.

The wiring prototype (Fig. 3a) was implemented as a more direct manipulation approach to end-user composition, compared to that of the filtered lists. Draggable nodes contained within a composition workspace represent the different components. Each node shows a symbol and the name of the component it represents.

To form a composition, a user first selects the trigger node and then drags a graphical arrow or "wire" to the intended responding node. The graphical arrows indicate the direction of control flow.

After two nodes have been connected, the user can specify the respective events and services that are to be part of the composition via dropdown lists contained within the nodes.

Existing compositions can be modified by rearranging the connecting arrows between nodes.



(a) (b)

Fig. 3. (a) Wiring diagram prototype. (b) Jigsaw puzzle prototype (with close-up)

Jigsaw Puzzle Prototype. The jigsaw puzzle metaphor is the alternative among the evaluated alternatives with the most direct mapping to a real-world object. Similar to the wiring prototype, this prototype also draws on the principle of direct manipulation. Users specify compositions by "snapping" together (dragging and dropping) two interlocking puzzle-like GUI elements representing various components (Fig. 3b).

Puzzle pieces with knobs represent trigger components, while puzzle pieces with holes represent response components. Similar to the wiring diagram, the trigger and response services are specified via dropdown lists.

For each composition involving a specific component, a new puzzle piece representing that component has to be collected from the "tray" (left), and dragged onto the composition workspace (right). Thus, the workspace can contain multiple pieces representing the same component (one per composition). This is in contrast to the wiring prototype, in which a single node is used for all compositions the corresponding component is to be part of.

Compositions can be deleted by separating two attached pieces.

4.3 Experimental Setup

Test setting and Equipment. The experiment was conducted in a usability laboratory equipped with high fidelity video and audio recording equipment. Screen mirroring software allowed us to record video of GUI interaction during the trials. A conventional PC with keyboard and mouse were used in the testing of the GUI prototypes. The prototypes were tested in random order.

To allow test participants to validate the run-time effect of compositions a software simulation tool was used. The simulation tool could give a simple visual representation of which and how components were affected given a specified condition.

Composition Tasks. The test participants were given a set of seven composition tasks (Fig. 4) to solve using the prototypes.

1. I want to be notified quickly in case of possible burglary or fire in my home.
2. The lighting in my bedroom should be on when I am there and off when I am not there.
3. I often take photos with my mobile phone when I am on travel. It would be nice if photos I am sending to my home could be displayed on the wall in the living room.
4. I am often in a hurry in the morning before I leave for work, but I usually need a cup of coffee to wake myself up. It would have been helpful if I could wake up to freshly brewed coffee.
5. I often wonder what the weather is like when I awake in the morning on working days.
6. When I am watching TV at home, I do not always hear if the doorbell rings.
7. I regret that I asked the house to brew coffee when I wake in the morning.

Fig. 4. Composition tasks

To avoid that the formulation of the composition tasks led to preference bias toward certain interfaces, tasks were presented as problems as opposed to instructions explicitly informing test subject what to compose. Thus, for many tasks there were multiple solutions. The motivation behind the problem-based tasks was to encourage interaction with the prototypes and to evoke creativity among participants.

5 Results and Analysis

5.1 Results

Filtered Lists. Nearly all test subjects stated that they experienced the filtered lists concept as a highly efficient way of specifying compositions. Although we did not measure task execution time, video recordings from the testing indicate that participants managed to come up with solutions to the composition tasks fairly rapidly compared to the other prototypes. Test subjects expressed that the interface provided a good overview over the different components and their associated services, and that it felt like a (quote) “safe”, “confident” and “logical” solution.

Many test subjects found that the orderly presentation of the components and services helped give them a quick idea of how to specify compositions. It was also pointed out that the GUI made it made easy to see alternative solutions to composition tasks.

Wiring Diagram. The filtered lists prototype gave explicit indications to users which screen elements that were intended for denoting the trigger part, and which screen objects that were intended for denoting the response part of a composition. The different parts were specified sequentially through lists organized in a left-to-right fashion. In the wiring diagram prototype, however, the issue of control flow between trigger and response component was not as apparent. Here, the users had to specify control flow by means of the connecting arrows. We noticed that participants, particularly those with no programming background, often did not pay attention to control flow when using the wiring diagram prototype. Their comments suggested that they did not think of the arrows as a way of specifying control flow, but rather a way of specifying two objects that should exchange data.

There were mixed opinions regarding how appropriate the wiring diagram prototype served as a composition interface. Some participants expressed that it was difficult to visually "trace" the connecting arrow to see which components it linked. Test subjects also expressed that as the number of compositions increased, the connecting arrows would clutter the interface, making it difficult to get an overview of existing compositions. One user commented that saved compositions could preferably be listed in a tabular form, as in the filtered lists prototype. Concerns related to overview of compositions arose primarily in the planning of new compositions. In this process, we found that test subjects frequently reviewed existing compositions to evaluate the how new composition would fit into the collection of existing ones.

There were also general concerns that the current approach would require users to frequently tidy up the user interface to get a clear overview of specified compositions. While the lack of overview was seen as a challenge in the process of specifying compositions, some test participants found that the presentation could be suitable for visualizing relationship between devices, for example, how many compositions a particular device is involved in.

Jigsaw Puzzle. The jigsaw puzzle prototype was generally viewed as an intuitive and fun way of building compositions. Participants commented that the interlocking pieces made it easy to see which devices that were part of a composition. The ability to provide this type of information clearly was one of the main challenges associated with the wiring diagram, where users often found themselves "tracing wires".

At the same time, it was frequently pointed out that the jigsaw puzzle prototype, similar to the case with the wiring diagram prototype, tended to gradually become disorderly and difficult to extract information from as the number of compositions increased. The number compositions that were specified during the task solving (typically between eight to ten) was, according to several test subjects, close to the limit of composition that could be visualized in the current GUI without significantly compromising the general overview.

We observed that some test subjects found it challenging to differ between trigger and response device when using the jigsaw puzzle prototype. After being presented a task, they would often put together pieces representing relevant components, but not pay attention to which piece that formed the trigger part and which piece that formed the response part. Similar issues related to control flow also applied to the wiring diagram prototype.

Test participants also gave indications that it was challenging to quickly understand and extract meaning out of compositions represented as jigsaws. Similar remarks related to difficulties in interpreting graphical representations of compositions quickly were also given in the testing of the wiring diagram prototype.

Test participants generally expressed that they enjoyed the direct manipulation of screen objects that the jigsaw prototype allowed for. The video recordings also indicate that the prototype encouraged users to try out alternative solutions to the composition tasks.

5.2 Analysis

Based on an investigation of the test subjects' transcribed responses, we identified four factors, or reoccurring areas of concern, that strongly affect the usability of the composition interfaces. These were (1) *predictability of composition model*, (2) *readability of composition representation,* (3) *overview and means for planning compositions*, and (4) *attractiveness and desirability*.

Predictability of Composition Model. The composition model used in the current study was equivalent to condition-response rules. While the three GUI prototypes all matched this model, comments form the test subjects and their interaction with the prototypes, suggested that the extent to which each solution effectively communicated the underlying composition model varied.

The user-perceived predictability (or lack thereof) of the composition model can in many ways be seen as a result of the way GUI elements for specifying compositions were arranged in the different prototypes. In the filtered lists prototype, the fixed left-to-right arrangement of GUI elements for specifying composition aligned with the format of condition-action rules. This was not the case for the other two prototypes. The nodes representing each device in the wiring diagram were initially distributed across the composition workspace. The jigsaw pieces were initially placed in the "tray" beside the composition workspace.

The user-perceived predictability of the composition model can also be seen in light of control issues that test participants experienced. Because the filtered lists prototype required the condition and response part to be specified in a sequential manner users avoided problems related to specification of control flow, which was sometimes the case with the other prototypes. With respect to the wiring diagram and jigsaw puzzle prototypes, control flow had to be handled explicitly by the users. For the wiring prototype this meant pointing the connecting arrows between nodes in the "right" direction. For the jigsaw puzzle control flow was specified in a left-to-right fashion (knobs and holes).

The aspects pointed out above, concerning arrangement of GUI elements and control flow, explain why test subjects experienced a feeling of greater control over the filtered lists prototype vis-à-vis the others.

Readability of Composition Representation. The usability of end-user composition interfaces does not only depend on their ability to provide intelligible interfaces for combining different components. We also found that test subjects frequently took existing compositions represented in the GUI in consideration when planning which devices and services that should be part of new compositions. This makes it essential that the user interface also effectively communicates the overall picture of compositions. One important lesson learned from the experiment is that what makes a potentially suitable metaphor for conveying to users how to manipulate GUI elements to form compositions, might in some cases compromise composition readability.

Our findings related to the wiring diagram and jigsaw puzzle prototype exemplifies how certain notations can have negative effects on readability of compositions. The representation of compositions in these two prototypes required more cognitive effort to "decode" and understand the run-time effect. In the case of the wiring diagram prototype this decoding meant that users had to trail wires to identify associated components, and also pay attention to pointing direction of the attached arrow. With regard to the jigsaw puzzle prototype, test subjects found themselves mentally "decomposing" puzzles.

Overview and Means for Planning Compositions. Another issue which surfaced as test participants were in the process of forming new compositions, relates to the means for planning compositions ahead. As described earlier, the wiring diagram prototype and the jigsaw puzzle prototype did not explicitly distinguish between the composition workspace and the spaces providing overview of existing compositions.. A reoccurring concern among test participants regarding these solutions was that the GUIs ability to effectively present a comprehensible overview of compositions was reduced as the number of compositions increased. In a sense, the user interface metaphor and the graphical elements used to represent compositions (wires and jigsaw puzzle pieces) would at a certain point have a negative effect on the overall overview.

We consider the issues concerning readability and overview to be particularly relevant for design of composition editors that allow for more complex compositions. Challenges related to, e.g., cognitive efforts involved in extracting meaning from composition representations could easily escalate in such cases.

Attractiveness and Desirability. The last factor, which we identified as being central for the user-perceived usability of composition techniques, relates to the overall experience of use. In both the wiring prototype and the jigsaw puzzle prototype compositions were formed by means of direct manipulation (drag and drop) of interface elements. Judging from the feedback from test participants, they experienced these solutions, particularly the latter, as being more playful and engaging compared to the filtered lists prototype. The reoccurring argument among test participants with a preference for the jigsaw puzzle prototype was that is offered a fun way of forming compositions. As described earlier, this approach also appeared to encourage exploration of alternative solutions to the tasks.

While most test subjects experienced the filtered lists prototype as being the most efficient in use among the alternatives, responses such as those described above indicate that designing appealing compositing interfaces is important for user acceptance. In more recent HCI literature, it has been argued that aesthetics and desirability in design are just as important as usability [7].

We are aware that the "fun factor" of the jigsaw prototype might have been given different priority depending on the consequences of making undesirable compositions.

This study was conducted in a laboratory setting and run-time effects of compositions were only simulated. In a real use environment, particularly in the cases related to home security, confidence in use might be given higher priority.

6 Concluding Remarks

We consider the identified usability factors to be essential parts of the "value set" against which end-user appraise composition interfaces. As such, we recommend that designers are mindful about these factors during design.

Acknowledgements. This work was funded in part by the ISIS project (Norwegian Research Council grant # 180122) and a joint research collaboration between SINTEF ICT and Telenor ASA. Thanks to MSc. Bjarne Muri for contributing to prototyping and usability experiments.

References

1. Maestre, J.M., Camacho, E.F.: Smart home interoperability: the DomoEsi project approach. International Journal of Smart Home 3(3) (2009)
2. ISO 9241-210: Ergonomics of human-system interaction – Part 210: Human- centred design for interactive systems. International Organization for Standardization (2010)
3. Humble, J., et al.: "Playing with the bits" user-configuration of ubiquitous domestic environments. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 256–263. Springer, Heidelberg (2003)
4. Newman, M.W.: Now We're Cooking: Recipes for End-User Service Composition in the Digital Home. In: CHI 2006 Workshop: IT@Home: Unraveling Complexities of Networked Devices in the Home (2006)
5. Sohn, T., Dey, A.: iCAP: an informal tool for interactive prototyping of context-aware applications. In: CHI 2003 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA. ACM, New York (2003)
6. Koskela, T., Väänänen-Vainio-Mattila, K.: Evolution towards smart home environments: empirical evaluation of three user interfaces. Personal and Ubiquitous Computing 8(3-4), 234–240 (2004)
7. Norman, D.A.: Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books, New York (2004)

Improving Code Reading and Comprehension on Large Displays

Selvihan Nazlı Kaptan[1] and Mehmet Göktürk[2]

[1] Bahçeşehir University, Çırağan Caddesi, 34353 Beşiktaş, İstanbul, Turkey
[2] Gebze Institute of Technology, Gebze, 41400, Kocaeli, Turkey
syavuzer@bahcesehir.edu.tr, gokturk@bilmuh.gyte.edu.tr

Abstract. Due to advances in display technologies and continuous decrease in large display prices, more users are choosing larger displays or multiple monitors for personal and professional use although standard size desktop monitors are still widely used. As programmers use a larger display surfaces to read and understand their code, current code editors are designed for standard monitor sizes and they do not exploit the extra space that comes with a larger display. In this paper, we discuss the use of a large display for code reading and test whether code reading can be improved by utilizing larger screen space.

Keywords: large displays, user performance, usability, code reading, reading and comprehension.

1 Introduction

Even though many programmers now prefer a larger display surface to read and understand code, one can see that current code editors are designed for standard monitor sizes. These code editors do not address advantages of extra space that comes with larger displays and multiple monitor extended desktops. One can understand that most lines of code is short in length and many programmers tend to split longer lines with multiple short ones as programming standards and practice advise shorter (approximately maximum of 80 characters due to historical reasons) line lengths.

Most programmers do not use extra large fonts even when they move to a large viewing space and they keep their previous user settings (font sizes, etc.). This results in a large code editor window with half of the code window being frequently empty. To address this issue, we propose to utilize the empty half of a large screen to provide user with textual information on the source code being manipulated aiming to reduce the mental load on the programmer.

In our preliminary survey, we tried to identify basic problems that programmers experience when reading a piece of code. To proceed on this track, it is quite important to know the type of information programmers search for most while navigating/scrolling since this knowledge can help to reduce navigational needs. When processing a single line of code, each variable declaration, function call or object creation is a potential reason for a navigational requirement as reader will need to recall a previous fact, and if not successful, he will have to navigate to the related

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 128–134, 2011.

part of code. According to our survey results, while reading a piece of software, participants spend time on navigating between lines to follow the code and memorizing functions/variables. Over 80% of the participants reported that they navigate to the same part of the code to recall information when they are reading source code. Furthermore, when the participants navigate back to the line where they left off, they re-read the same lines to remember (e.g. after navigating to a function definition, and scrolling back to the line where the function was called). Navigating back and forth, short-term memory is used inefficiently and the reader consumes more mental resources. This also relates to the age related problems that occur in programmers where coding ability usually decrease after certain age.

2 Related Work

Research on reading electronic text traditionally suggests reading on a computer screen is slower than reading from a book; however, no significant effect on comprehension is observed. Moreover, reading speed is affected by the format of information that is the number of characters per line, the number of lines per page and interlines spacing affects the speed at which one reads [1]. This suggests that reading speed will be different on larger displays where more lines will fit into a single page.

Studies on large displays were mostly focused on collaboration, and less has been done to discover the benefits to individual users. Richardson et al (1989) studied the effects of display size on readers' comprehension and manipulation on electronic texts. User experiments showed comprehension rate and performance did not change with display size, yet users preferred larger displays for reading [2]. On the other hand, Lin and Hsieh (1996) measured the effects of physical display size on task completion time as a part of their user experiments, and in result they concluded users completed tasks faster on larger displays while this did not change the correct response percentage.

Code reading, the process of careful reading of source code for modification, refinement and debugging purposes, has been one of the most important skills to work in both industries and academia [3]. Reading process involves both resolution of information and decoding the location of information in 2D space [4]. Thus, users form a cognitive map that defines where a word or a phrase resides within the text. The task of reading and understanding a source code mostly requires re-reading certain part of it. Cognitive mapping activity is expected to be higher when reading a piece of code since the reader will try to create a map of variables and values, functions and algorithms in order to understand the code being viewed.

This process of mapping can be thought as encoding the source document. In regular texts, highlighting has been considered as an effective tool in encoding process. Research has repeatedly shown that tactics such as highlighting, note taking, and underlining help with the encoding process. These tactics on computers has been found to decrease the time needed to locate a target [5]. This may suggest marking a piece of information as important decreases the time spent on reviewing the relevant material to access it.

3 User Experiment

Goal behind this study is to reveal the differences between users' code comprehension performances on medium and large displays. For code assistance window, textual views of code is preferred because text has the advantages of begin easily communicated, effectively manipulated and highly scalable. A textual model may also improve comprehension by expressing information directly within the source code being manipulated. Any attempt to successfully observe the effects of code assistance tool on a large display should address the following issues: (i) On large displays, more lines will be visible at a time. This may provide a better performance on larger displays than on medium displays. This difference should be distinguished from the improvement introduced by the code assistance tool. (ii) Many studies have shown that even with similar backgrounds, performances of programmers show significant differences [6, 7, and 8]. Personality is also a factor which affects the programmers' performance on different stages of development process.

3.1 Physical Experiment Setup

In the experiments, 2 (two) different screens have been used. One 17" Acer AL1716 and one 40" Alba LCD TV. In each experiment users are given a display, a mouse, a pen and paper-based task assignment. Since users are not expected to type, keyboards and other peripherals were removed from the setup.

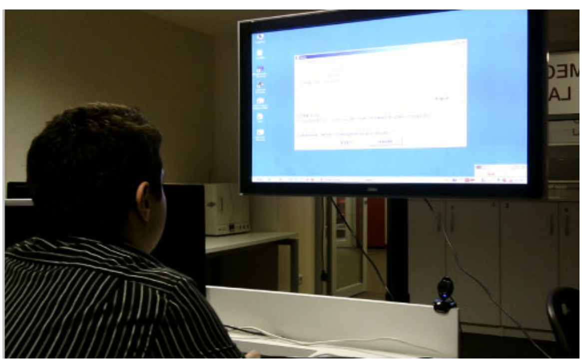


Fig. 1. Experiment setup

Displays are placed to achieve a FOV between 45°-60° which is the range where users performed better in numerous studies. Participants were free to adjust their positions as they wish in order to simulate a comfortable working environment.

Participants were recorded during the entire experiment. For each experiment, response time, number of correct responses and scroll displacements were also logged.

3.2 Subjects

Subjects were sophomore and junior students of Department of Computer Science at Bahçeşehir University in a course on Data Structures and Algorithms (using C++). In total, 24 students participated and all students had attended at least two computer science classes before.

3.3 Objects

Objects are simple C++ programs that use basic programming structures such as arrays, functions, variable declarations, logic/arithmetic expressions and input/output commands. To minimize performance differences between participants based on their level of knowledge in C++, programs used in experiments are kept as simple as possible.

3.4 Experiment Procedure

Each experiment consisted of an output evaluation for a given C++ program. Each subject experienced all three setups and was time-tested with 2 experiments on each display setup. For each display setup, participants were placed in the lab and were given a C++ code and they were asked to determine its output. Experiment timer started as soon as the subject accepts to see the code to be evaluated.

The experiments were conducted in three main groups; medium display size with no code assistance, large display with no code assistance, large display with code assistance. Code-assistance window is populated with textual information on the code being evaluated. Code-assistance window contents were statically prepared for each source code separately and loaded from files with the assignment code. This information window code included reminders on variable declarations, function declarations and previous and further references to these (according to currently active line of code). To simplify the generation of window contents, source codes for the tasks are divided into main segments, and for each segment block of text information is prepared. When the participants focused on a line (when the cursor is active on a line) within a segment, segment-related part of the text information was loaded in the assistance window.

After completing two experiments on a display, users scored their assignment according to their difficulty on 0-5 scale where 5 indicates hardest and 0 the easiest task.

4 Results

Response times collected during user experiments are normalized and are given in Figure 2. While there is slight improvement on response times between medium display and large display with no code assistance, this difference needs to be explored in detail with further experiments. Response times showed that code comprehension performance is higher with code assistance window used on large displays.

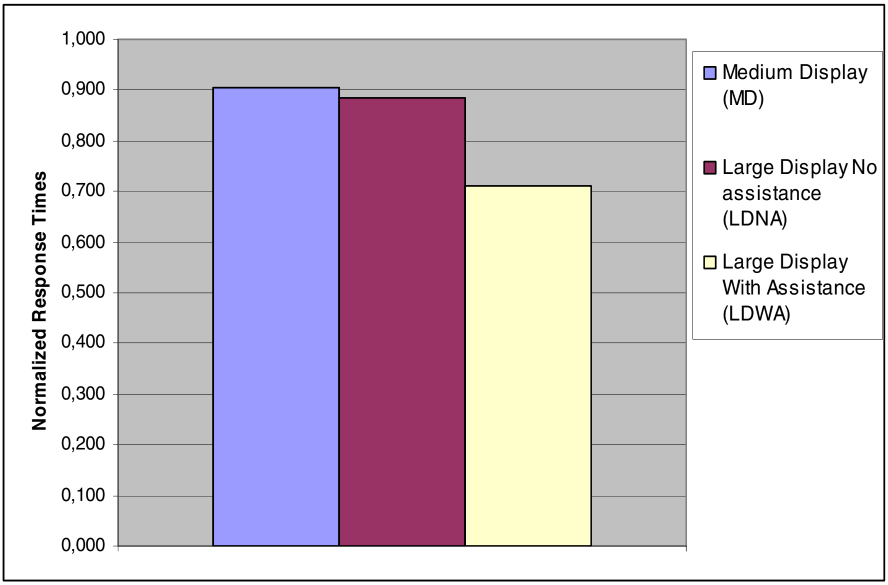


Fig. 2. Normalized response times on three different setups

After the experiments on each setup, participants were asked to score the task difficulties. Figure 3 summarizes subjects' responses to task difficulties. The differences between 3 setups are important since tasks are similar and monitor sizes and code assistance is altered. Figure shows subjects consider code comprehension as more difficult on medium displays then on large displays.

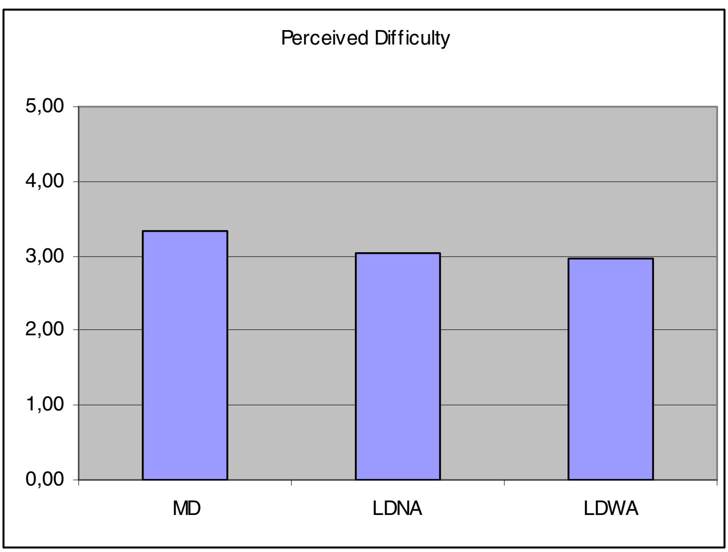


Fig. 3. Users comments on task difficulties on different setups

Although response times benefit from code assistance window, perceived task difficulty scores are close on large displays with and without code assistance frame. On the other hand, the difference between medium displays and large displays is mostly based on reduced navigation requirements.

5 Conclusions

In this study we aimed to examine how the skills of participants in understanding the functioning of a piece of code are affected by reading it on different display sizes and with and without code-reading assistance window. In user experiments, we used a code-reading assistance frame located on the empty half of the large display which included textual information on the code being manipulated. The window presents constant reminders on specific code parts (variables and function declarations of interest and previous references to these), recently visited code blocks and comments extracted from the current code block.

Our results suggest that task completion times are improved by the use of larger display space and reading assistance frame, although success rates are not significantly affected by the setup. Although majority of programmers still use displays less than 22” in size, larger displays, dual (or multiple) monitors and projection systems are becoming prevalent. As a common skill in both industries and academia, both professional programmers and programming course students may benefit from extended display surface with the help of appropriate assistance tools which do not exist in currently available environments.

References

1. Kruk, R.S., Muter, P.: Reading of Continuous Text on Video Screens. Human Factors 26(3), 339–345 (1984)
2. Richardson, J., Dillon, A., McKnight, C.: The effect of window size on reading and manipulating electronic text. In: Megaw, E. (ed.) Contemporary Ergonomics 1989, pp. 474–479. Taylor and Francis, London (1989)
3. Ueno, D.: Linenum- Info: System Support for Code Reading. In: Eighth IEEE International Conference on Advanced Learning Technologies, ICALT 2008, July 1-5, pp. 600–602 (2008)
4. Wästlund, E., Norlander, T., Archer, T.: The effect of page layout on mental workload: A dual-task experiment. Comput. Hum. Behav. 24(3), 1229–1245 (2008)
5. Fisher, L., Coury, B.G., Tengs, T.O., Duffey, S.A.: Minimizing the time to search visual displays: the role of highlighting. Human Factors 31(2), 167–182 (1989)
6. Shneiderman, B.: Software psychology: human factors in computer and information systems. Winthrop Publishers, Cambridge (1980)
7. Grant, E.E., Sackman, H.: An Exploratory Investigation of Programmer Performance Under On-Line and Off-Line Conditions. IEEE Transactions on Human Factors in Electronics HFE-8(1), 33–48 (1967)
8. Boehm, B.W.: Software engineering economics. Prentice-Hall, Englewood Cliffs (1981)

9. Goldberg, J., Helfman, J.: Evaluating User Expectations for Widescreen Content Layout. Paper presented at the Usability Professionals' Association Conference, Austin, TX, USA (June 2007)
10. Tan, D.S., Gergle, D., Scupelli, P., Pausch, R.: Physically large displays improve performance on spatial tasks. ACM Trans. Comput.-Hum. Interact. 13(1), 71–99 (2006)
11. Kawashima, H., Oda, K., Kobayashi, A., Suzuki, M.: Does limited window size affect text comprehension? In: Vision 2005 - Proceedings of the International Congress held between 4 and 7 April 2005, London, UK. International Congress Series, vol. 1282, pp. 622–626 (September 2005) ISSN 0531-5131

Designing the AR Experience: Tools and Tips for Mobile Augmented Reality UX Design

Gini Keating, Daniel Guest, Anne Konertz, Niccolo Padovani, and Andrea Villa

Qualcomm, Corporate Research and Development,
5775 Morehouse Drive, San Diego, California, USA
{GKeating,DGuest,AKonertz,NiccoloP,Avilla}@Qualcomm.com

Abstract. User experience design professionals have established tools and processes for creating user interfaces. However some new interaction paradigms differ so greatly from their predecessors that established methodology is inadequate. Mobile augmented reality (AR) is one such paradigm. Viewing and interacting with 3D registered graphics that are overlaid in the user's camera view creates novel design challenges.

This paper describes the shortcomings of current user centered design methodology when applied to the creation of user experiences for mobile AR. We address the challenges of designing a seamless 3D mix of real-world and virtual experiences along with the complexity arising from the location-based nature of the content.

Finally, new tools and methodologies are presented for researching, designing, rapid prototyping, and user testing of mobile AR user experiences, with the intent of helping designers overcome the unique challenges inherent to this new interaction paradigm.

Keywords: User Experience, User Centered Design, Human-Centered Design, Mobile Augmented Reality, Rapid Prototyping, Paper Prototyping, User Testing.

1 Introduction

Since the formal introduction of design processes focused on the user [1], human-centered design has grown into standard practice [2]. Also referred to as user centered design (UCD) [3], it often includes various types of user research, iterative design, rapid prototyping, and user testing. [4]

In this paper we show how we expand upon the current tools and methodology in order to design experiences for mobile AR.

We will not address AR technologies, but instead will provide pre-tech and technology-agnostic user experience (UX) tools practitioners can apply to aid in the design of mobile AR experiences.

1.1 Mobile Augmented Reality Definition

For the purpose of this paper, we will use the following definition of augmented reality [5]:

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 135–141, 2011.

1) Combines real and virtual.
2) Is interactive in real time.
3) Is registered in 3D.

In mobile AR, “real” is the camera view, and “virtual” is overlaid augmentation graphics.

For mobile AR in this paper, we add one more component:

4) Is accessed using a fully mobile device (e.g. cellular phone, tablet, etc.).

Thus the user can experience AR anywhere, enabling interaction in any environment.

1.2 Impact on User Centered Design

When designing mobile experiences for AR, we found the biggest challenges to be:

1) Camera view as a user interface (UI) element: Instead of building on a blank screen, the base of the mobile AR UI is the camera’s view of the surrounding environment.
2) Environment location: What appears in the camera view is generally location specific. Additionally the overlay augmentation content can be location specific or location agnostic depending on the design. We will deal primarily with location specific cases as they offer the most UCD challenges.
3) 3D graphics registered in a 3D environment: The augmentation graphics need to be simultaneously integrated into their environment and yet draw the users attention to important informational content.

In the following sections, we put forth the process and tools we used to address these challenges in pre-tech user experience design.

2 Rapid No-Tech Prototypes

Traditionally rapid paper prototyping is a tool for quickly mocking up user interfaces for design exploration and user testing [6]. Designers draw the interface on paper, often drawing one screen per page or adding sticky notes to enable switching between states. This is a powerful tool for exploring ideas before the technical implementation starts, allowing for rapid iterations with zero engineering impact.

When designing for mobile AR, we wanted to use paper prototyping to explore experiences before starting 3D model creation and technical implementation. Thorough definition and review of graphic direction is especially important in AR compared to traditional 2D UI designs because 3D graphic model creation and animation is notably more time consuming. Consequently, any change in design after the 3D model creation has begun can have a significantly negative impact on budget and schedule.

The challenges around 3D graphics apply not only for mobile AR but also for any 3D UI environment (e.g. gaming). However, in AR there is the added challenge of the graphics being registered in 3D over the camera view. The following steps show

how we adjusted the rapid prototyping process to overcome these challenges for mobile AR.

When we applied traditional paper prototyping to AR, it quickly became apparent that a key part of the interface, the camera view, was missing. When designing interfaces for AR, we have additional requirements for imitating the camera view from different perspectives, envisioning a real scene from all possible angles, and taking environmental conditions into account.

To add the camera view, we first created a pass-through paper prototype. In this version, a rectangle the size of the screen was removed from the paper so the real world seen through the hole became the camera view. Next, sticky notes placed at the edges could intrude on the view to represent overlaid information. This worked well in creating a 2D UI layer in front of the camera view but did not fulfil the need to emulate the augmentation graphics aligned to the environment.

Our second step was to place the augmentation data in the real world by adhering sticky notes (representing the data) on top of objects around the room, then observing the scene through the pass-through paper prototype. This worked well to simulate 2D data registered in 3D space, such as star ratings augmentation attached to a book. But this second step fell short for prototyping 3D graphic augmentation.

To simulate the 3D graphics, our third step was to create physical 3D models by folding paper tab-and-slot style to roughly represent 3D computer generated models. When they are placed in the environment then observed through a pass-through paper prototype, the simulation is complete. The user can move around the environment, viewing the mock augmentation from any angle.

This 3D paper prototype rapidly showed its worth. The team was designing a research application that included a virtual dolphin sticking its head out of virtual water and interacting with the user. In the 2D paper prototype version, we used a series of sticky notes depicting the animation of the dolphin. When viewed through the pass-though prototype, the experience seemed promising. Then we made a 3D paper mock-up of the dolphin and the water. It quickly became clear that we had not chosen the optimal animal or environment for the experience. When looking at the 3D dolphin model through the pass-though prototype, the user could move to low viewing angles where he or she could see under the virtual water to the truncated bottom of the dolphin in a rather disturbing way. We changed the dolphin in water to a penguin on land, so that the animal image would always appear whole and intact. The rapid prototyping process had worked as no computer modelling or coding had started, and therefore the change had minor impact to schedule and resources.

The pass-through paper prototype along with 3D paper models have become a mainstay in our process to quickly check ideas for viability.

3 Environment Prototypes

The content of the real-time camera view is an integral part of the mobile AR experience. Thus, it was necessary for the UX team to conduct field experiments in many different locations in order to design for the wide variety of backgrounds that may appear in the camera view.

For example, to research a mobile AR application that displays restaurant information, it was useful to observe locations close to several restaurants before beginning to design, then revisit the locations during the design process. Yet it was difficult to rapidly iterate a design when each change meant a new field visit to check the impact in situ. Consequently, we developed two tools for prototyping real-world environments in the lab to simulate remote locations: 1) a small-scale street model and 2) a full-scale panorama.

The first tool is a simple 3D street model. For example, we recreated a busy urban street in 1:12 scale. Buildings were created using photos taped to paper cubes. Environmental elements such as cars and pedestrians were added to the scene by using scale model toys. This small-scale street model allowed us to pre-plan content for the real environment, rout pedestrian paths and check sightlines. This enabled us to be more productive in the office and better prepared on our field visits.

Moving forward with a design, the full project team (user experience, engineers, and business functions) made multiple field visits to the chosen environment. The trips were very productive, as we could all stand together in situ to discuss the experience. However, to better enable collaboration and to speed design and testing, we found it desirable to recreate this in situ experience in the lab.

Thus, we constructed a full-scale panorama of our chosen environment: a busy urban street. First we erected a 180-degree arch framework with a seven-foot radius and a height of seven feet. Then, we applied a large-scale panoramic photographic that was taken in the chosen location. Standing at the focal point, we recreated the user's view so that it was nearly identical to being on the busy urban street. This full-scale panorama worked well as a replacement environment for the majority of the future "field visits" (now lab visits).

When the first round of technical prototyping began, we found we could test the on-device experience in the lab by orientating the panorama in the same compass position as the real-world environment, and then hard-set the mobile device's GPS information to match the coordinates of where the photograph was taken. Thusly aligned, when a user stood at the focal point of the full-scale panorama, the on-device augmentation experience in lab very closely mimicked how the application would look in the actual environment. The results were that 1) the content in the camera view when looking at the panorama was nearly flawless, and 2) the presentation and alignment of the virtual augmentation matched as well. The full-scale panorama enabled us to rapidly iterate and test in the lab before taking users out to the actual environment.

4 User Testing

Formal user testing by specified users in a specified context is a standard part of the UCD process [7]. When the specified context is not location-specific, the testing can be done in the controlled environment of a usability lab. However, for mobile AR, we wanted to test users in the real-world environments where they might use the applications. This requirement added challenges to planning and executing tests. As an example, we will discuss our Balboa Park exploratory user testing (hereafter referred to as "the park test").

Early employment of exploratory user testing helps UX designers understand the user's mental model of the new experience while simultaneously noting the effectiveness of preliminary design concepts [8]. For exploratory user testing of mobile AR, one of the main tasks we chose to test was exploring local content. We sought meaningful data collection with minimal impact on each users time (90 minutes). Thus we needed an engaging real-world environment that was rich with diverse content that our users could explore freely. We chose Balboa Park (San Diego, California, USA) because it has museums, restaurants, shops and gardens all within walking distance.

Compared to a lab test, preparing for the park test was notably more elaborate. No controlling or changing of the environment of the park was possible. Over the roughly .15 square mile area, we prepared for variables such as: weather, pedestrian density, traffic, street performers, etc. We made multiple trips to the park to scout the area, take pictures and determine appropriate augmentation content to seed our proposed AR database.

Accumulating knowledge about the park ahead of time prepared the moderator to effectively discourage users away from tangents that would take them out of the data rich areas, or into areas where they might not feel comfortable. When testing in such a large exploratory area, we found that it is necessary to have adequate preparation to make the most of the available time. In the lab, if the user takes a tangent unrelated to the research question at hand, this generally emerges in the form of spoken thought, and therefore is relatively easy to recover from. However, during the mobile AR test, such tangents were more significant as they took the user off of their current walking path, and thus took more time to recover from.

Determining the best data capturing methodology for the park test took some pre-research and experimentation. We prefer to have 3 forms of data from user testing: 1) a capture of the onscreen interactions, 2) a video of the user, and 3) the researchers' notes.

In the lab, this is achieved with 1) cameras fixed on the ceiling to pick up the onscreen interactions, 2) cameras on the walls to capture the user's facial expressions and comments, and 3) researchers and observers behind the glass taking notes.

To mimic this in the field, we determined the best data capture could be obtained as followed. 1) Automated screen capture software on the device recording the interface plus the user interactions. 2) A trailing videographer with a wireless microphone on the user to record his or her statements, actions and environment. 3) A moderator in close proximity to the user, providing general guidance and asking questions. A comprehensive debriefing interview containing follow-up questions helped address the lack of note taking during the test.

Later, during data evaluation, the screen capture feed and environment video feed were manually synced, allowing us to better understand user intent, delight, confusion, etc. in relation to what visuals the user saw on the screen and in the real world. As project team observation during the test was not practical at the park, the detailed processing and presentation of the data for this test was crucial.

As a result of the detailed pre-planning and modified data capturing, the park test yielded robust data from the field, which equaled the high quality we expect from in-lab testing.

5 Content Design

The inclusion of the camera view as a UI element presented challenges when designing mobile content for AR. Through experimentation, we determined different styles were appropriate for different types of applications.

We first experimented with photo realistic augmentation that closely matched the visual style of the camera view. For example, we placed a realistic virtual banner as augmentation on a real-world building. The result had very little initial visual impact, as most users could not discern what content has been added. The user had to closely compare the real-world view to the onscreen view as a "search for the difference" task. When the users discovered the banner, they were surprised and engaged in the illusion. While photo realistic augmentation was rewarding as a novel experience, this visual style would not be appropriate when a designer wants to direct the user's attention for tasks such as information sharing and discovery.

For 2D informational content such as hours on a store, bright designs with dark outlines stand out against diverse backgrounds and draw the user's attention.

For 3D discovery content such as a dinosaur in front of a distant museum, cartooned graphic styles combined with animation draw attention in contrast to the real-world background.

For more immersive experiences, such as tabletop gaming, or closely exploring the above-mentioned dinosaur, the graphic styles can be much more diverse as the imagery clearly takes over larger portions of the camera view.

Thus, many diverse graphic styles can be utilized in creating augmentation content for mobile AR, as long as the style is inline with the intended use.

6 Conclusion

As a new interaction paradigm, mobile AR enables new user experiences, which come with new challenges. UCD processes need to grow and change to address these challenges.

From the earliest prototypes through to the last round of user testing, designers should mindfully consider the intricacy of designing a seamless 3D mix of real-world and virtual experiences along with the complexity arising from the location-based nature of the content.

Acknowledgments. The authors would like to acknowledge the following members of Qualcomm, Corporate Research and Development, for their contributions: Joel Bernarte, Sebastien Mounier, Ken Olson.

References

1. Norman, D.: The Design of Everyday Things. Doubleday, New York (1988) ISBN: 978-0465067107
2. International Standards Organization: 13407 Human-Centered Design Processes for Interactive Systems. In: ISO/IEC 13407:1999 (1999)

3. Abras, C., Maloney-Krichmar, D., Preece, J.: User-Centered Design. In: Bainbridge, W. (ed.) Encyclopedia of Human-Computer Interaction. Sage Publications, Thousand Oaks (2004)
4. Mao, J., Vredenburg, K., Smith, P., Carey, T.: The State of User-Centered Design Practice. Communications of the ACM 48(3) (2005)
5. Azuma, R.: A Survey of Augmented Reality. In: Presence: Teleoperators and Virtual Environments, pp. 355–385 (August 1997)
6. Virzi, R., Sokolov, J., Karis, D.: Usability Problem Identification Using Both Low- and High-Fidelity Prototypes. In: Common Ground: CHI 1996 Proceedings of the SIG CHI Conference on Human Factors in Computing Systems. ACM, New York (1996) ISBN:0-89791-777-4
7. International Standards Organization: 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. In ISO 9241-11:1998 (1998)
8. Rubin, J.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, pp. 31–33. John Wiley & Sons, Inc., New York (1994) ISBN: 0-471-59403-2

Dra*Wii*ng Together: Exploring Collaborative User Engagement in Art Exhibitions

Hyungsin Kim[1], Hyun Jean Lee[2], and Ellen Yi-Luen Do[1,3]

[1] GVU Center & School of Interactive Computing
[2] The Graduate School of Communication and Arts, Yonsei University
134 Shinchon-Dong, Seodaemun-Gu, Seoul, Republic of Korea
[3]College of Architecture,
Georgia Institute of Technology, Atlanta, Georgia 30332, USA
{Hyungsin,ellendo}@gatech.edu,
hyunjean@yonsei.ac.kr

Abstract. There is growing interest in user experience studies in Human Computer Interaction (HCI) and Interaction Design. Many researchers focus on designing technology to enhance user experience, specifically for engagement, joy, and collaboration. In order to explore user engagement, we developed three different WiiArts applications: ***Illumination***, ***RippleCast***, and ***ChromaFlow***. We performed data analysis based on the video data collected from three different art exhibitions in three different countries: the USA, Germany, and South Korea. In this paper, we present the results of our observations that identified users' engagement time, the number of people in a drawing collaboration session, and their drawing patterns. Then, we discuss the design implications for user engagement in terms of interactivity, collaboration, and creation. We conclude that both situated interaction and collaborative creation should be considered for designing technology for enhancing user engagement.

Keywords: Art Installation, User Engagement, Experience Design, Collaborative Interaction, Situated Interaction, Drawing Application, Media Arts, Interactive, Tangible Interaction, Wii Remote-based Physical Interaction.

1 Introduction

We live in a world in which technology is seamlessly embedded in our everyday lives. Automatic lighting systems in hallways, self-checkout systems in grocery stores, and wireless Internet connections throughout cities are good illustrations as to how our lives are deeply interconnected to technology. The notion of who are technology consumers has also changed. Besides tech-savvy and gadget-loving consumers, everyday people with no technical background are buying and using technology. Without knowing the logic behind the development, people enjoy using the technologies if they give them fun and joyful experiences. Furthermore, this notion pushes Human Computer Interaction (HCI) researchers to explore user experience and engagement with technology [1, 2]. Traditionally, HCI research focuses on the usability and accessibility of design technology in terms of efficiency

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 142–151, 2011.

and effectiveness [1]. Usability is the most important factor to consider during the system evaluation stage because technology is usually designed to perform certain specific tasks [3, 4]. For example, HCI researchers usually focus their attentions to issues such as whether the system was designed well enough for users to be able to create documents, websites, and photo albums as they wish, or whether the button was visible and in the right location for the tasks. These are all important matters for research. However, with the extended needs and use of technology in our everyday lives, it is also important to consider user experience factors in designing technology. The question is then, how can we design technology to enhance user experience? In another word, when designing technology, how can we encourage users to actively engage in the experience? Greenburg and Buxton stated the importance of user engagement in their paper. They criticized the current practice of HCI and argued that usability evaluation could be harmful in many instances [5].

In this paper, we explored the design factors for enhancing user engagement. For research purposes, we first developed three different WiiArts applications: ***Illumination***, ***ChromaFlow***, and ***RippleCast*** to explore user engagement. Then in order to collect real-life use data, we conducted observations in actual gallery art exhibitions in three different countries: the United States of America, Germany, and South Korea. Based on direct in situ and videotaped observations, we analyzed the data and decided on the factors for enhancing user engagement.

2 Interactive Arts

In *Art as Experience* (1984), John Dewey says that a work of art is an individualized participating experience [6]. A work of art is recreated every time it is aesthetically experienced by the viewer. Each viewer creates an imaginative relationship with the self through his/her experience with artwork, and this kind of process can be referred to as "interactive engagement." In this sense, all artwork is interactive. If we could agree that to a certain degree, all artwork is open to participation and multiple interpretations by viewers, depending on their own experiences with the artwork. Why then, do we call some types of artwork interactive art, and how does this kind of interactive art differ from the other art forms that involve interactive experience?

The goal of interactive art is to involve the spectator in some way.[1] Interactive art is a type of art that involves a viewer's active participation. Unlike the traditional art piece, such as a painting or sculpture, interactive art encourages viewers to create their own narratives [7]. Interactive artists often claim that their work is open to the audience's participation [8]. Therefore, their work has an open-ended structure that allows or even encourages the audience to "complete" their creations. In this sense, with respect to interactive artwork, the relationship between the creator and the audience becomes blurred, sometimes to the extent that some may even say that the

[1] Itsuo Sakane, the Japanese journalist and curator, suggests that interactive art is simply art that involves the participation of the viewer. However, he goes on to remark, "All arts can be called interactive in a deep sense if we consider viewing and interpreting a work of art as a kind of participation" (1989, 3).

audience becomes a co-creator of the work [9].[2] Therefore, user engagement is a critical component in designing an interactive art system [7].

As for its system, new media art or interactive art not only utilizes new media (computer-based media) as its medium for artwork creation, but also exploits the interaction between humans and computers for a new way to experience art. Since most interactive artwork uses the computing system as a communicative means, the interactive artwork often focuses on how to make a reliable input system that can sense the participant's interaction and present a complicated output that can provide new experiences, which are possible only with computer-based creation. Although this kind of interactive artwork does not explicitly aim for functionality and usability in its interaction with the system (in the sense that the system requires an interactive and communicative process between humans and computers), sometimes the focus is still not very much different from that of the HCI type of computing system design. Despite the fact that there are shared interests between HCI research and interactive art, this interactive art as new media art must have its own interests and directions as a genre of art that distinguishes it from commercially applied scientific and technological research.

If the interactive experience is intended as artwork, it can have different goals. We can imagine that there will be some differences with regard to the subjective reading in such experiences. Although some works encourage this more, whereas other works encourage this less, artists may want to explore, question, and challenge the relationship between the viewer and the interactive system. They may want to elicit diverse and flexible dialogues between the work and the viewer in order to expand the scope of the experience to evoke a dialogue between the self and the world beyond. In "The Construction of Experience: Interface as Content (1998)[3]," interactive artist David Rokeby mentions that in interactive art, responsibility is at the heart of interactivity as the system's ability to respond to its interactors' interaction. Therefore, the process of designing an interaction should also, in and of itself, be interactive. In designing the interface, interactive artists need to pay "close attention to the user's responses and make modifications as a result of our observations [10]." In this way, interactive artists need to "expand the terms of this interactive feedback loop from simply measuring functionality and effectiveness, to include an awareness of the impressions an interaction leaves on the user and the ways these impressions

[2] The Antennae Of The Race by Nina Colosi, Producer/Curator, evo1
http://www.evo1.org/essays.html
"Audience: The networked, digital environment is by nature polyvocal and favors a plurality of discourses. Interactive art involves reciprocity and collaboration between the creator or creators, the audience and a project. Audiences collaborate in the process of remapping textual, visual, kinetic and aural components of the artwork—the public and audience becomes a participant.
Artist: Rather than being the creator of a work of art, the artist often becomes a mediatory agent and facilitator for audiences' interaction with and contribution to the artwork."

[3] David Rokeby,, "The Construction of Experience: Interface as Content"
http://homepage.mac.com/davidrokeby/experience.html. This article appears in the book: "Digital Illusion: Entertaining the Future with High Technology," Clark Dodsworth, Jr., Contributing Editor 1998 by the ACM Press, a division of the Association for Computing Machinery, Inc. (ACM) published by Addison-Wesley Publishing Company.

change the user's experience of the world [10]." As a result, he emphasizes that the interface can be a content of interactive works. Interactive artists need to seek profound and subtle ways that the interface itself shapes our experience of that content by defining how we perceive and navigate content.

We also think that in this participation and interaction paradigm, the most important thing is not only interacting with the self, but also engaging in collaboration through interaction. The viewer may himself/herself find different meanings in the piece on different days, at different hours, or at different stages of his/her own development. Furthermore, it is important for artists to well understand how users interact with their work because audience engagement is the one of the most critical factors in public galleries [11]. In fact, the setting where interactive artwork is exhibited in the gallery and appreciated by the viewers affects the interactive experience of the work. The dilemma of an interactive work is how to guide its interaction without specifically or visibly noting so next to the work. Sometimes without a note, there is hardly any guidance to the audience in terms of how to interact with the work. With a note, however, it can delimit the audience's free and voluntary response to the work. In this situation, in the gallery, other audience members' actions and reactions bring about a tremendous effect to others' interactions with the work since people's curiosity about a piece of work is affected by observing other people's interactions and reactions. They get clues from others about how to interact with the work. Thus in interactive artwork, unlike other non-interactive artwork, audience engagement brings into the exhibition arena certain kinds of collaboration with other viewers.

3 The WiiArts Applications

The purpose of the WiiArts applications is to let users, or viewers create experimental video, audio, and image processing art projects through collaborative and expressive experience. The applications are developed in Max/MSP/Jitter and Java using pre-existing sensing technologies, such as Nintendo's Wii Remotes and a wireless sensor bar [12]. To explore user engagement through interactive art systems, we chose the Wii Remote as the interaction method for several reasons. First, the Wii Remote can provide an environment where multiple users can play together in front of a large projection screen. Second, it can allow users to interact with and manipulate items on a screen via movement or pointing, due to its motion-sensing capability through accelerometers and infrared detection. In this section, we will describe the three applications respectively: ***Illumination***, ***RippleCast***, and ***ChromaFlow***.

3.1 Illumination

Illumination lets participants draw anything with real-time fluid candlelight traces. The projection screen becomes a shared drawing canvas, and up to three audience members can draw simultaneously each with their own Wii Remotes. The candlelight source imagery of the three burning candles is captured by three cameras in real-time. Thus, the three different candlelight traces drawn by three participants can be composed together to create a dynamic drawing. The goal of the interactivity is to give participants contemplative aesthetic experiences by drawing with candlelight in a dark space. Figure 1 shows six different drawings created in the exhibitions.

Fig. 1. Illumination

The drawing located at the top right corner of Figure 1 shows a creation of the word "WiiArt" created by two participants. Since candlelight drawing does not continue more than 5-seconds, after several trial-and-error attempts, the two participants have to work collaboratively to complete the word together within the time constraint. After succeeding in the first collaborative writing, the participants continued to create other objects, such as a house (Figure 1 center right screen), and a heart shape (Figure 1 bottom right screen).

3.2 RippleCast and RippleCast 2009

RippleCast attempts to provide a calm engagement and a chance to encounter nature in an interactive media art experience. As if the participants were standing on the shore of a lake, the projected image of a still pond on a big gallery wall lies waiting for the viewers. A viewer may take hold of one of the controllers and start to move it around as players do in Wii games and watch the ripples magically appear on the tranquil lake scene in front of them (Figure 2).

In *RippleCast*, interaction with the Wii Remote is designed to mimic the movement of stone skipping like we do with our hands. Similar to the way we grip a real stone with our fingers, stretch our arms and release the stone from our fingers at the last moment, the participant can trace a curved line in the air while holding the Wii Remote and pressing the big button on the bottom with his/her index finger, and finally release the button to release a virtual stone. As a result, ripples form on the surface of the still pond, depending on where the participant throws the stone. Based on the strength and degree of the throwing motion, the stone skips, making one or more hops. Certain movements, directions, and speeds are programmed to get better results. Tracking of the movement and speed of the Wii Remote are measured with its embedded accelerometer, and the data are wirelessly transmitted to a computer in a separate room via Bluetooth technology.

Fig. 2. RippleCast 2009

After the success with *RippleCast*, we developed *RippleCast 2009* to include wavering water in a pond rather than a still image. *RippleCast 2009* was implemented with a real-time video of lightly wavering water pond to create a more engaging experience. Figure 2 above shows three participants competing with one another to see who could throw a virtual stone further.

3.3 ChromaFlow

Like *RippleCast*, *ChromaFlow* is a WiiArts application that draws ripples on the screen. The interaction and program use exactly the same methodology as *RippleCast*, but while *RippleCast* emphasizes self-reflection through the experience, *ChromaFlow* instead promotes a more collaborative interaction among users. In this work, each person's Wii Remote draws single colored ripples. When each stroke with the Wii Remote is made and a virtual water drop is cast on the surface, it bounces several times as in *RippleCast*, like a splash. Currently, this application supports interaction for up to three people at a time. Thus, if three participants work together, they can cast ripples in three different colors from three Wii Remotes. The ripples cast by the participants flow on the surface of the screen. When they spread and diffuse near other ripples, they start to mix, creating a colorful image, as seen in Figure 3.

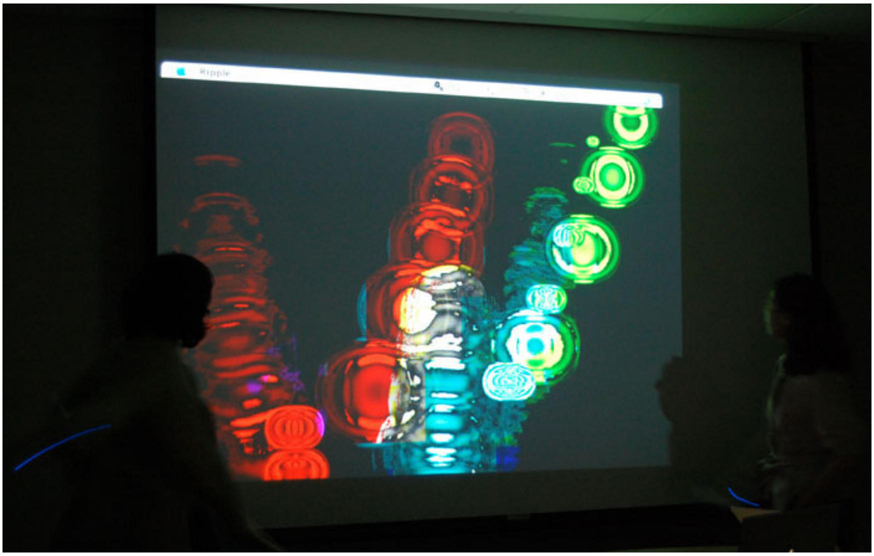


Fig. 3. ChromaFlow

Sometimes ripples are dropped at unexpected locations, much like occasional random drops in watercolor drawings. This kind of effect can bring about unpredictable results created only by chance. *ChromaFlow* is reminiscent to one of Jackson Pollock's drip paintings in the sense that the audience can draw abstract images with ripples through diverse and active body gestures [13]. Unlike Pollock's canvas, which was placed on the floor, the canvas screen in *ChromaFlow* is set up vertically on the wall. Also in *ChromaFlow*, the shared canvas becomes an open field for the creation of many different kinds of abstract visual works. Similarly to *RippleCast*, the position of where ripples drop on the surface are mapped from the Wii Remote interaction in a way that is unpredictable. This is because the gestures with the Wii Remotes are made in open space; thus, every time a movement is made, its direction or position is made a little bit differently. With the waving colorful water surface on the screen, the audience becomes immersed in a calm and reflective mode.

4 Design Implications

In this section, we first describe the summary of our observations in terms of users' engagement time, the number of people in collaboration, and their drawing patterns. Then, we propose that user engagement should consider two factors: situated interaction and collaborative creation. The exhibitions were held in three different countries: Seoul, Korea; Atlanta, Georgia, USA; and Bonn, Germany, respectively. The first two exhibitions were parts of the author's solo and group exhibitions. The last one in Germany was a part of the second ACM Tangible Embedded and Embodied Interaction Conference Demo session [12]. Due to the characteristics of the exhibition, most of the audience members were adults.

Our user engagement analysis focused on the duration, number of people, and drawing patterns. We tried to answer three research questions: (1) How much time did the audience spends playing the WiiArts application? (2) How many people (up to a maximum of three users) played the WiiArts application together? and (3) Did the audience create specific meaningful drawing patterns?

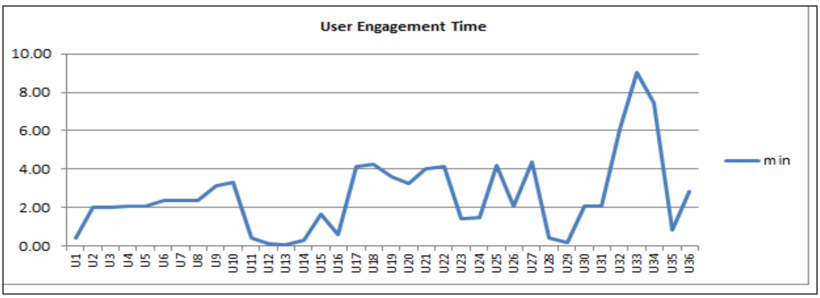


Fig. 4. User Engagement Time

Figure 4 shows the graph of an individual user's engagement time. A total of 36 participants played with the WiiArts applications. Overall, the time people spent on the applications is diverse (min: 27 seconds and max: 9.03 minutes). Some people played in pairs or groups of three. Some of them would play alone after collaborating with others to further explore the interactivity and its unexpected outcomes. Since we do not have enough data to co-relate each application with its playing-time, we did not conduct more in-depth playing-time analysis. However, the overall time shows that people really enjoyed using the applications – the participants spent an average of 2.6 minutes interacting with the applications in the public exhibitions.

Our observation analysis clearly shows that three attributes such as interactivity, collaboration, and creation can increase user engagement in a public art exhibition. We frame the attributes by calling them **"situated interaction"** and **"collaborative creation."**

First, **situated interaction** can be defined as situations in which people would engage more when the interaction is truly situated within specific contexts. For example, let's take a look at the *RippleCast* application. When the participants first see the screen, they may not understand the purpose of the image of the lake. Since it is displayed in a gallery, they may assume that it is an artwork as part of an exhibition. They may also recognize that it does not seem to match with the semiotic reading of art photography or art videos. Yet, still they may feel liberated looking at the imagery of the pond, which transports them into the feel of being in nature, in the countryside. From the moment they find three Wii Remote controllers with bright blue LED lights in a corner of the room, they start to discover a connection between the controllers and the natural scene. They may not get a response immediately, since the interaction requires that they press a specific button to interact with the image. Nevertheless, occasionally some of them would be able to, either by luck, by accident, or by exploration, create a small ripple reflection on the surface of the pond. At that moment, they suddenly realize that they can skip stones over the still pond, as if they were actually in front of water in nature. Thus, the interaction (interacting with the system) is truly situated. Comparing the participants' feedback between *RippleCast* and *RippleCast 2009* also shows the importance of contexts. Participants who played *RippleCast 2009* were more engaged by spending more time because the lightly wavering water in a pond enabled participants to feel more situated in the context.

Second, **collaborative creation** can be defined as the positive relationship between the activity of co-creating and the levels of engagement. We observed that most people played the application as a pair or in groups of three. As expected, participants spent more time when they were collaborating with others to draw something with *Illumination* or competing by throwing his or her stone to see which one would go further using *RippleCast*. In addition, how a group of participants used the shared space (a big canvas in this case) also illustrated the importance of collaborative creation. For example, in *Illumination* application when one participant drew a line from the top-left corner, another participant drew some object in the middle. Then the third participant used the top-right space to draw his own star. Seamlessly, the one big-shared screen was smoothly utilized by the group. The collaborative creations facilitate smooth interaction and encourage positive user engagement.

5 Conclusions and Future Direction

This paper explored how users, especially exhibition visitors, interact with the WiiArts interactive art applications in three different places. Based on our qualitative and quantitative analyses, we concluded that audience engagement is intricately co-related to situated interaction, as well as collaborative creation. This work derives from investigating user engagement through the lens of interactive media art. Our study can shed lights on designing a system to support non-task-based work, such as an explorative system. For future work, in order to gain a more in-depth understanding of the co-relations among users, engagement, and systems, we plan to deploy the WiiArts Applications for a longitudinal study in a public exhibition.

Acknowledgments. We thank all of the exhibition visitors who participated in playing the WiiArts applications and giving us their precious feedback. We thank Gaurav Gupta, who helped create the initial stage of WiiArts applications and Vinutha Prabhakar, who helped code the initial video observation. Lastly, we thank all of the art organizers in the United States, Korea, and Germany who invited us for their exhibitions.

References

1. Dix, A., Finlay, J.E., Abowd, G.D., Beale, R.: Human-Computer Interaction. Prentice Hall, Englewood Cliffs (2003)
2. Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Hum.-Comput. Stud. 64, 79–102 (2006)
3. Sharp, H., Rogers, Y., Preece, J.: Interaction Design: Beyond Human-Computer Interaction. Wiley, Chichester (2007)
4. Dourish, P.: Where the Action Is: The Foundations of Embodied Interaction. The MIT Press, Cambridge (2004)
5. Greenberg, S., Buxton, B.: Usability evaluation considered harmful (some of the time). In: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 111–120. ACM, New York (2008)
6. Dewey, J.: Art as Experience. Perigee Trade (1980)
7. Sengers, P., Gaver, B.: Staying open to interpretation: engaging multiple meanings in design and evaluation. In: Proceedings of the 6th Conference on Designing Interactive Systems, pp. 99–108. ACM, New York (2006)
8. Peters, C., Castellano, G., de Freitas, S.: An exploration of user engagement in HCI. In: Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, pp. 9:1–9:3. ACM, New York (2009)
9. Hansen, M.B.: New Philosophy for New Media. The MIT Press, Cambridge (2004)
10. Rokeby, D.: The construction of experience: interface as content. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1988)
11. Sengers, P., McCarthy, J., Dourish, P.: Reflective HCI: articulating an agenda for critical practice. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 1683–1686. ACM, New York (2006)
12. Lee, H., Kim, H., Gupta, G., Mazalek, A.: WiiArts: creating collaborative art experience with WiiRemote interaction. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, pp. 33–36. ACM, New York (2008)
13. Pollock, J.: Lavender Mist: Number 1, Painting (1950)

Versatile Wearable Computer for Drivers

Gyouhyung Kyung[1], Songyi Chae[1], Kyung Hyun Nam[1], Kyungmin Lee[2], and Wanjae Shin[2]

[1] School of Design and Human Engineering, Ulsan National Institute of Science and Technology, Korea
[2]Human Factors & Devices Research Team, Hyundai-Kia Motors, Korea
{ghkyung,robim0314,skarudgus}@unist.ac.kr, {kmlee,downer2k}@hyundai.com

Abstract. A versatile wearable computer for drivers (VWCD) was proposed that can extend in-vehicle multi-modal display spaces. In the first development phase, LEDs, mini video displays, earphones, and vibrators were included as visual, auditory and tactile displays, and were all attached to an eyeglass frame. One electrocardiographic electrode was attached to the VWCD to obtain the driver's heart rate signal at the posterior auricular artery, and a gyro sensor was used to track the driver's head position. Finally, a Bluetooth device was included to enable communication between VWCD and its mobile phone platform.

Keywords: multi-modal display, visual display, wearable computer for drivers.

1 Introduction

One of the major reasons to use a wearable computer is that it can seamlessly provide information to people on the move. While wearable computers have been used in various application areas including games, medical devices, aerospace, and automotive industry (for maintenance), there are some problems in current wearable computers that hinder their commercial success. In terms of design and human factors aspects, many wearable computers are still obtrusive and uncomfortable due to their bulky and heavy components such as control box and cables. Therefore, size minimization and wireless connection are a key factor for improving its design. In addition, many wearable devices still tend to be expensive and inefficient [1]. Currently, there is no commercialized wearable computer for drivers [2], despite surface vehicles are most commonly used to move people and objects.

Among display types, visual display is most widely used in the vehicle, followed by auditory and tactile displays. Visual information in the vehicle is usually displayed in the area placed in front of the driver (e.g., cluster, crash pad, and windshield), which is however not always included in the driver's forward field of view (e.g., when the driver look at the outside mirror or the passenger on his side). If the driver does not look ahead, it is very likely for him to miss critical information displayed on conventional in-vehicle visual displays. Therefore, it would be very effective to track the driver's head position and provide critical information right in front of the driver's eye. By doing so, the driver would be less likely to miss

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 152–155, 2011.

important information, even when the driver does not look at the forward direction of the vehicle.

There are an ever increasing number of drivers who wear eyeglasses, due to the rapidly growing aging driver population with degraded eyesight. Wearable devices which are mounted on, or integrated into, the eyeglasses, hence, seem to a promising solution to provide safety-related information for drivers. VWCD has been conceptualized to continuously provide the driver safety-related information, as well as to monitor driver's health information. By integrating LEDs, speakers and vibrators on the frame of glasses as visual, auditory, and tactile displays, the driver can always be informed of the direction of imminent risky objects and situations, regardless of their head position, unless their eyes are closed. With electrocardiogram sensors which are attached to the leg ends of the eyeglasses, VWCD can record the driver's heart rate at the bilateral posterior auricular arteries.

VWCD has been designed with the aims to minimize its size and weight, and to consider its appearance, and connection with widely used nomadic devices (e.g., iPhone and iPod Touch). It should also improve the driver's situational awareness.

2 Related Research

To implement wearable computers, a variety of display technologies have been developed together. There are various types of display technologies such as single crystal silicon, liquid crystal on silicon, high temperature poly silicon, low temperature poly silicon, laser or optical beam-scan, organic light-emitting diode, digital micro-mirror device, and micro-cathode ray tubes. For cost effectiveness and minimization, VWCD uses LEDs and mini video displays to provide visual information to both eyes.

Currently, there are five different methods to display visual information; binocular immersive, binocular see-periphery, binocular see-through (navigation/video sunglasses), monocular see-around, and monocular see-through. As a first phase of developing VWCD, binocular see-around types and mini video display have been selected to provide additional visual information, while minimizing obstruction of

Table 1. Ambient vs. Focal Visual Systems (from [6])

	Ambient System	Focal System
Primary Functions	Visual guidance; motor control	Form recognition; identification
LGN (Lateral Geniculate Nucleus) source	Magnocellular	Parvocellular
Cortical stream	Dorsal stream	Ventral stream
Spatial resolution	Low	High
Contrast sensitivity	Asymptotic at low (10%) contrast	Requires mid-to-high contrast
Spatial frame of reference	Egocentric (absolute body coordinates)	Allocentric (relative object space)
Temporal resolution	High	Low
Primary control mode	Closed-loop	Open-loop
Memory requirements	Low	Moderate-highs

focal vision [6]. Ambient and focal visual systems have different properties, and hence these differences should be considered when designing a visual display system for drivers.

There are several commercialized eyeglasses-type wearable computers such as a wearable display developed by Konica Minolta, an image overlay system for medical data visualization, the Olympus Mobile Eye-Trek model, and the Personal Display by Eyewear.

3 VWCD System Development

3.1 Innovative System Concept Generation by TRIZ and Lateral Thinking Tools

In order to provide drivers with context-driven information, collecting appropriate and accurate information from the driver and the vehicle is essential. One of the best ways is to use physiological data such as brain waves, heart rate, and body temperature, while minimizing the driver's distraction. Systems that measure drivers' fatigue (e.g., MIT smart car [3]) have not been successfully commercialized due to intrusiveness of the system.

According to Edward De Bono, the problem solver should explore different ways of examining a challenging task [4]. By adopting his concept, called Lateral Thinking, we can think of a system that can obtain the driver's physiological data as well as provide relevant information to the driver. As a measuring device, an eyeglass-type system appears to be less intrusive than other types of alternatives for drivers (clothes, wrist-band type, etc), considering that many people already get used to wearing corrective eyeglasses or sun glasses. In addition, VWCD provides drivers with information necessary for safe driving while allowing them to maintain their forward sight. Furthermore, multi-modal (visual, tactile, and auditory) displays enhance information transfer by redundancy gain and provide selective, effective display channels, depending on driving context. The multi-functionality and dimensionality change suggested by TRIZ to develop inventive systems are hence applied to the VWCD concept generation, resulting that it functions as input as well as output device, and extends space for visual and auditory displays and sensors to eyeglasses [5].

3.2 System Components

The VWCD system consists of a pair of see-around visual displays, mini video displays, vibrators, earphones, and electrocardiogram sensors, a gyro sensor, an operating system, two control / processing units, and a Bluetooth transceiver. Fourteen LEDs are attached around each lens (Figures 1 and 2). Gyro sensor was used to detect rotation and direction of the driver's head movement. LEDs were controlled by the control unit and when drivers turn their head to the right direction while the potentially risky object is detected from the front or left direction, then the left vibrator turns on and the left LEDs blink to alert the driver. When the driver looks at his right side, with the head rotated, left LEDs and vibrators come into play. LEDs can communicate wirelessly with a laptop via the Bluetooth module. Visual displays and earphones can be connected to a mobile application, iPod Touch, and the driver can see the icon such as traffic signals using their peripheral vision and hear the alerting sound. Two electrocardiogram sensors are connected to the processing unit and these continuously record the driver's heart activities.

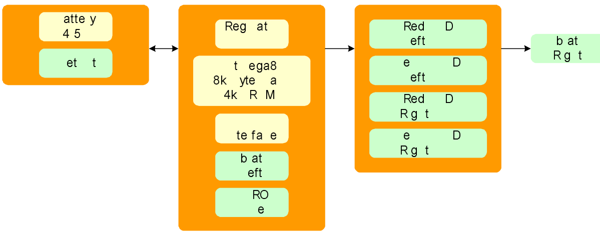


Fig. 1. VWCD block diagram

Fig. 2. VWCD prototype

4 Future Work

Visually overloaded drivers can easily miss important information. VWCD extends display design space, and facilitates critical information reception by providing multi-modal information. However, it should be determined later what combination of input modalities is optimal for the drivers. In the current development phase, a see-around visual display type was selected majorly due to its low cost and functional simplicity, whereas an optical see-through type has not been considered due to its high cost and technological immaturity. In the future, visual displays such as video and optical see-through might be used to more effectively detect the forward driving situation and to display critical information directly on lenses during driving. VWCD always contacts with the driver's skin, which is necessary condition to continuously monitor health-related information. VWCD and its processing unit are wire connected. As a fashionable product integrated with new technologies, VWCD components should be wirelessly connected and miniaturized further to increase users' technology acceptance, and hence, the possibility of commercial success.

References

1. Rügge, I., Ruthenbeck, C., Reiter, B.S.: Changes of HCI Methods towards the Development Process of Wearable Computing Solutions. In: Kurosu, M. (ed.) HCII 2009. LNCS, vol. 5619, pp. 302–311. Springer, Heidelberg (2009)
2. Randell, C.: Wearable Computing: A Review (2005)
3. Healey, J., Picard, R.: SmartCar: detecting driver stress. In: Proc.15th Int. Conf. Pattern Recognition, Barcelona, Spain, vol. 4, pp. 218–221 (2000)
4. De Bono, E.: New think; the use of lateral thinking in the generation of new ideas. Basic Books, New York (1968)
5. Altshuller, G.: 40 Principles Extended Edition: TRIZ Keys to Technical Innovation, vol. 1 (2005)
6. Castro, C.: Human Factors of Visual and Cognitive Performance in Driving. CRC Press, Boca Raton (2009)

Dynamic Navigation System Design for Networked Electric Vehicles

Frazer McKimm[1], Manuela Galli[2], and Veronica Cimolin[2]

[1] DHS, Shamrock Chambers 1-2 Eustace Street Dublin 2 Ireland
[2] Dipartimento di Bioingegneria, Politecnico di Milano, Milan, Italy

Abstract. Data saturation of satellite navigation systems (already a problem with location based services) will become particularly acute in the emerging area of networked electric vehicles (NEV). Sophisticated energy management and navigation software may solve a technology integration challenge, but it will leave unresolved the usability implications for drivers and fleet operators. These include navigation data specific to commercial electric vehicles; delivery scheduling, routes, times, traffic congestion avoidance, range & charge levels etc. Many are time dependent factors that complicate interaction with a map based navigation system. They also risk augmenting driver stress and distraction induced errors. This Paper has two objectives. Firstly we examine the problem of information saturation of navigation systems. Secondly we undertook a series of user tests to evaluate an alternative NEV navigation system. The DHS solution is a compressed data feed delivering "just in time" multimodal prompts embedded in the map route. The test results demonstrated improved driver comprehension and reduced driver glance away time from road to navigation system.

1 Research

Driving is a multi-tasking activity that requires drivers to manage their attention between various driving and non-driving-related tasks. "New studies show that drivers overestimate their own ability to safely multitask, even as they worry about the dangers of others doing it, device makers and auto companies acknowledge the risks of multitasking behind the wheel" [1]. This is especially the case in the use of in-car satellite navigation systems.

More significant is the way map information is presented on a satellite navigation system. These are based upon a 300 year old Ordinance Survey, grid reference maps system. Cartographic representations of an area are even older and more deeply embedded in our subconscious. The designers of these maps assumed the viewer was static (standing or sitting). Driving a vehicle presents quite a different view of reality, one which is dynamic and changing. If one adds dense context based information, then the static map model presents usability problems.

In understanding how we interpret visual information (like Sat Nav maps in a car) we need to understand how the human brain responds to visual stimuli in a dynamic environment. The brain has evolved to interpret a reality generally based upon a ***near***

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 156–166, 2011.

static or slowly changing environment. Even in the case of "3D" driver view maps, like the one illustrated below, (Fig. 1) the map is an externalized representation of a reality. It is ***not*** how the driver perceives the environment and the navigation "options" as they present themselves.

Fig. 1. Route as driver experiences it combines time & space

For modern men and women, running or rapid movement (outside the context of sport) is a relatively unnatural state of the brain. Decisions and precise calculations are made in this dynamic situation; however in such a state we are fully engaged with our sensory abilities and sensitive to a multitude of surface sounds and external stimuli. From an anthropological or historical perspective the running state was a normal one for primitive man, and the sedentary state of the modern mind is a recent one. Our minds are a summation of a very long evolutionary past and as such, less influenced by recent "modern" history. Carol V. Ward notes; "because the human mind and behavior are products of evolution, we must reconstruct the selective pressures that shaped our lineage in order to understand ourselves today." [2]

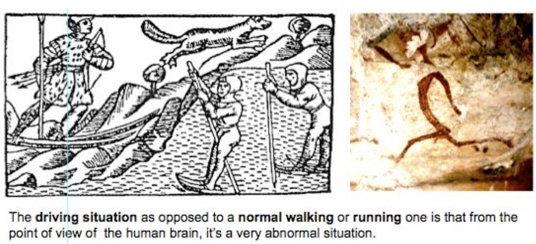


Fig. 2. Running is part of a normal mind & body experience

The driving experience in terms of the brain's interpretative abilities can thus be considered an even more abnormal state for us, as our body is static and sensory detection is visually centric. From inside the car, the driver's view or perspective appears as a movement down a "tunnel" like space with a precise time relation.

From an evolutionary perspective, driving is a paradox for the mind. We are rushing towards a target at speed, but our body is static and at the same time senses we are moving. There are a number of ways for the driver to view this; we are flying through the air horizontally or we are in a sort of simulated running. This running or rapid movement is to the brain an alert situation, one more related to our primeval **“hunter state”**.

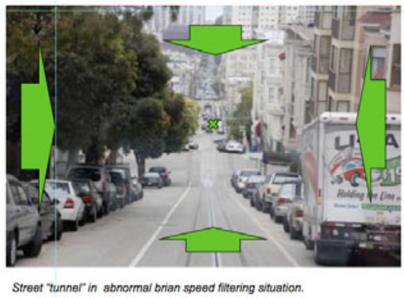


Fig. 3. In driving rapidly through a space anticipation is strong

For the brain, "stress response—also known as the fight-or-flight response—involves the release of stress hormones like **adrenaline**… High levels of adrenaline were meant to serve as emergency fuel generators in times of existential danger, such as when our forefathers were being chased by a wooly mammoth.." [3]

Our theory is that this rapid movement in the car is closer to the “hunter state” (and with its increased adrenaline levels) is the only “similar” condition we have to the driving experience. Drivers of all ages cope with this differently; however, the addition of densely packed data via the navigation system will act to increase apprehension on the part of the driver. Driving along a road forces the mind to make split second adjustments. In this state the body is effectively on alert, the brain is keenly focused on anticipating what is coming towards the driver. Regarding this heightened adrenaline state, much research has been conducted in relation to the “time dilation effect “ or the perception of extended time due to heightened levels of adrenaline in the body *e*specially in threatening situations *[4]*

This offers the interesting postulation that a user now can comprehend and achieve far more in a split second of navigation interaction then in a more relaxed or less adrenaline filled state. We don’t agree that a solution to driver distraction caused by data-rich satellite navigation systems lies within this line of reasoning. Research in the area of augmented reality maps where adding “a virtual object with proper inclusions in the real scene” [5] would indicate a viable future solution, but costs are currently too high for mass application.

The visually centric nature of the information in the dynamic car state presents an interpretative dilemma for the brain. Humans have evolved to understand a dynamic running-based state and engage all our senses to interpret the environment in real time in order to make and /or anticipate events affecting our safety. Adrenaline release serves to engender a physical reaction, (flight or fight) one that cannot be done while seated in a car. Secondly, the body’s own production of adrenaline is designed for short “abnormal” periods and de facto extended adrenaline periods should not be a standard side effect of driving. Thirdly, we are physically neutered, (seated and still) with a truncated sensory capacity, depending primarily on the visual faculty. We are also multi-tasking while in motion: driving, navigating and /or talking on the phone. Whereas the hunter /athlete in us can display great ambidextrous skill while in motion, less can be said of driving and multi-tasking. Driving and talking on the phone, for example, produces a level of cognitive awareness and dexterity analogous with being drunk [6].

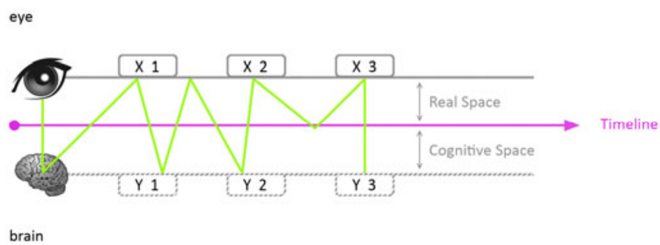


Fig. 4. The brain creates a cognitive version of the time line event sequence

A map view visual model does not take sufficiently into consideration the dynamic, time pressured nature of the car environment. The brain will compare what it is experiencing with what it knows. It has to rely primarily on the visual faculty to "read" both the real and the map representation of events. This does not help the driver in a quasi sensory deprived space to correctly interpret and react to a situation. The subject of sensory deprivation and its various forms has been exhaustively studied since the 1969 John Zubec report " Sensory Deprivation- Fifteen Years of Research". Peter Suedfeld in his introduction noted "perceptual deterioration" as one of many side effects. Though there is a significant difference between the Zubec experiments and a car cabin, parallels exist and relate to the importance of full sensory engagement in reacting to four principal causes of driver distraction and risks of accidents: **Visual** e.g., focusing on something other than the road, **Audible** e.g., someone talking. **Physica**l e.g., eating, **Cognitive** e.g., something that requires you to think about something other than driving. [12] It is reasonable then to state that due to the data rich and dynamic nature of fleet EV navigations systems, driver distraction and incomprehension will increase beyond that of existing navigation systems. This suggests that it is necessary to re-think how an interactive NEV navigation system is designed. Because the driver moves ***with*** the car and observes an instantaneous sequence of events, information must be structured and sequenced to match this dynamic. There is (the author believes) a sensory deprivation induced disconnect between an abstracted map model and the lived experience which the brain is normally equipped to deal with. Gallese and Lakoff in their "*interactionist* theory of meaning" identify the "same neural substrate" or connection between the conceptual and the living experience. "Conceptual knowledge is embodied, that is, it is mapped within our sensory-motor system." [12]. A data feed mechanism is needed to maintain a link between a dynamic but truncated sensory environment and the navigation map. The driver needs context specific, "just in time", multimodal packets of information. As we will discuss later, these "information packets" take the form of landmarks, some iconic or standard, others personal and inserted by the driver. Orientation signifiers are central, we believe, to an adaptive and more user-internalized navigation system. As stated earlier, if standard navigation systems are already a usability concern, then Networked Electric Vehicles represent a real challenge.

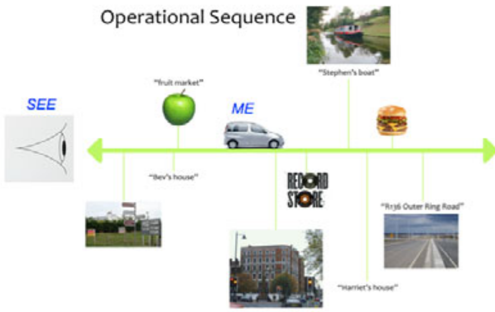


Fig. 5. Above a time line version of event sequence

Networked Electric Vehicles for Fleet use

A Networked Electric Vehicle is connected to an energy management backend which structures charging of the entire fleet as well as route selection according to electricity grid modulation and energy cost parameters. In the case of standard navigation systems, drivers generally have a single end destination. In the case of fleet NEV's there are multiple destinations and data specific to commercial electric vehicles; delivery scheduling, routes, times, traffic congestion avoidance, range & charge levels etc. Drivers at the best of times have a limited capacity to manage navigation information in the short time frame of a dynamic environment. This NEV data saturation complicates matters and represents another layer of dynamic information reaching the driver. With electric vehicles, charge point information and range distance will also be factored in.

Observing the above considerations, DHS looked to the sequence and structure of information in its research. The above diagram is a time-line representation of a navigation route where the driver has to reach or pass a number of landmarks. Landmarks have an added value in relation to time management and orientation in a delivery context. They aid a visual and auditory restructuring of the data feed as the characteristics of any in car information presentation should be recognizable, dynamic and compressed. Though a map view of the entire delivery routes would help before driving begins, a compressed "just in time" data presentation would, DHS believes, simplify the navigating experience and reduce driver distraction. In published literature on driver safety there have been numerous studies documenting the effects of distracting tasks while driving. In general, these studies demostrated that distracted drivers have slower responses to critical traffic events and are more likely to miss external events such as changing trafic light, among other effects ([9]*; [14]; [7]; [6]*) Many studies are also dedicated to the evaluation of the distraction effects related to the use of different kinds of navigation systems. [10]. In particular the results showed that the condition without voice required a higher demand on visual attention.

The addition of voice substantially reduced the visual attention requirements. These results are reached considering mainly the driver's eye glance behavior in terms of frequency and duration. [11]. Starting from the literature in this topic and from the limits present in current navigation systems, the DHS proposal took into consideration specific time and event stimuli.

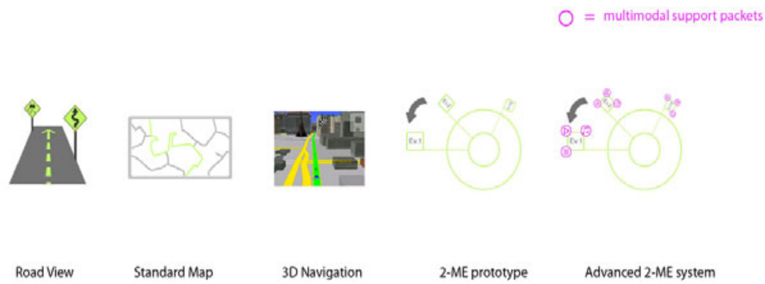


Fig. 6. Above 2-ME, early model for just in time information presentation

The aim of the DHS testing work was to evaluate how a dynamic, "just in time", restructured display (termed "2-ME") improves user comprehension, reduce distraction and ultimately increases driver safety. The display uses reconizable "land-marks" which act as prompts where route change decisions need to be taken. These data feeds are supported by auditory prompts located directionally (left, right, straight ahead) to further impove comprenension via multi modal command outputs. [12]

TESTING BACKGROUND: In order to evaluate its "Dynamic Navigation" theory, DHS engaged Dr Galli and Dr Cimolin of the Department of Bioengineering at the Politechnico di Milano, Italy, to design a series of tests which we hoped would prove that this system offered a viable alternative to traditional navigation systems. These tests are covered more comprehensively in a separate paper by the authors "Navigation With Adjusted Multi Modal Data Delivery," whereas the aim here is to explore the application of the concept to Networked Electric Fleet Vehicles.

TESTING METHOD: the drivers followed an urban route in a driving simulator answering to the voice commands given to them by three different navigation systems: the first Satellite navigation traditional system (Sat-Nav), the second (DHS version) based on the use of landmarks able to provide detailed information about the map (2-Me system), and the third (DHS version) with the same landmarks but with a different map representation (C-Me). The systems were evaluated in terms of the effects on driver distraction. In particular an optoelectronic system for kinematics analysis of head driver movements was used. The number of head movements to the navigator display and the time spent in watching the navigator were key evaluation parameters. In this way researchers were able to identify any distraction occurring in the different navigator modalities.

RESULTS: The 2-Me system is characterized by a low number of head movements to the navigator display and by less time spent by the drivers to read and understand the navigator map presented on it.

CONCLUSION: The evaluated navigation interface 2-Me shows promising elements for a new navigator interface planning.

APPLICATION: The messages provided by the 2-Me interface are not currently supplied by existing navigation systems. They could possibly be added to them in the near future.

MATERIAL & METHODS: The effects produced on driving by 3 interface navigation solutions (Sat-Nav, 2-Me interface and C – Me interface) were tested in this work. The Sat-Nav interface is the Satellite navigator system proposed actually by the market (Figure 7). The 2-Me and C-Me interfaces developed by DHS are characterized by the use of specific landmarks to indicate the way that the driver has to follow. Instead of the command proposed by the Sat-Nav system “after 140 m turn right”, in the 2-Me and C-Me modality the command is substituted by “at the church turn right”. In this way specific landmarks, related to the landscape, are used in order to better identify the street changes required by the driver. The selection of specific landmarks, related to the urban landscape perceptible as the easiest way by the driver, were chosen in order to answer the compressed information requirements. The two modalities (2-Me and C-Me) differ to each other as described in the following way: **2-Me** essentially presents information (to me the driver) in segments that arrives at the driver as he or she drives down a given road (figure 8). **C-Me** effectively gives the driver an over view (see me – the route) of the trip in one “flash” (figure 9). It connects the memory icons (drawings, photos of places of people) in a continuous strip. We structured our 2-Me and C-Me to allow for the 4-Dimensional Awareness model of the driver being in the car, with data elements arriving in sequence at him/her in a time space “tunnel” situation. We also used pattern recognition and abstraction scale to construct our 2-Me and C-Me solution.

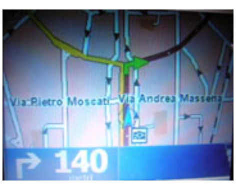


Fig. 7. Sat-Nav interface

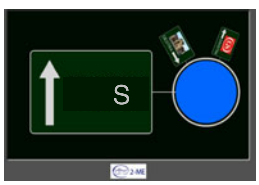


Fig. 8. 2-Me interface

Fig. 9. C-Me interface

20 subjects (11 males; 9 females, mean age 34 +/- 12 years) used to a navigation system driving were selected for the tests. The subjects were asked to drive in a simulated car environment, when in front of them a video taken from a real driving experience was presented. In particular the subjects were tested in these tasks:

Task 1- go to the destination following Sat -Nav system.
Task 2- go to the destination following 2-Me interface system.
Task 3- go to the destination following C-Me interface system.

In all the navigation interfaces the given command was both visual and vocal. The aim of this experimental set-up was to evidence how many times and how long the drivers looked at the navigator device in order to better understand the map. The subjects were analyzed during the simulated driving, analyzing their biomechanical behaviour using the 3D quantitative kinematic evaluation. In particular the effect produced by the 3 different interface was evaluated analyzing: 1) the number of the head movement corresponding to the number of time in which the driver look to the navigator display 2) the time spent to maintain the position of the head versus the navigation display corresponding to the time spent by the driver looking to the navigator display instead of the road. An optoelectronic system (Elite Bts it) able to measure the 3D coordinates (x, y, z) of passive marker was used for the test.

To know the head movement number to the navigator during the three different tasks , the angle α, in the horizontal plane, indentified by the central markers placed on the head and the x axis (corresponding to the anterior direction), was computed (figure 11). When α is =0° it means that the subject is looking to the road in front of him, when α shows a maximum as shown in figure 12, it means that the subject is looking to the navigator display. This behavior is of particular importance because the more time spent in the evaluation of the screen, the more distraction that is produced by the navigator interface. Starting for the angle α versus time the number and duration of head movements were considered:

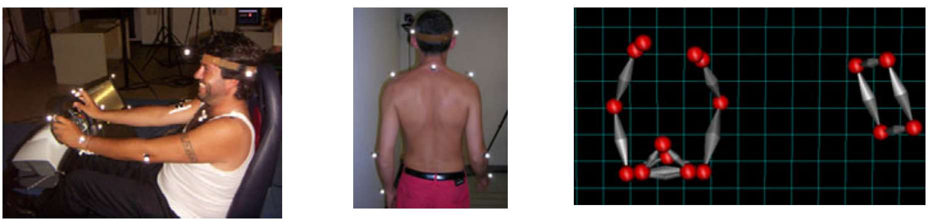
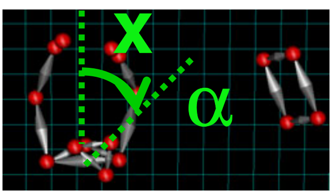


Fig. 10. Marker position (lateral view, back view) and 3D reconstruction (top view during driving)



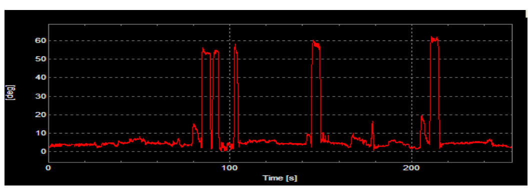


Fig. 11. Identification of angle □ for the head movement analysis **Fig. 12.** angles □ versus time. Peak values correspond to the position of the driver face to the navigator. The plateau duration indicates the time spent by the driver analyzing the navigator screen.

Two kinds of questionnaires were proposed to the drivers. The first one (based on 19 questions) was used to test if the drivers are or not experienced with navigator system driving and with electronic equipments in general. These data are useful in order to establish if the tested subjects could be considered as a homogeneous group. The second one (summarized in Table 1) was proposed in order to know the opinion of the drivers about the use of the three different interfaces. For each question of this questionnaire a score from 1 to 5 (1 not satisfied- 5 fully satisfied) was asked in order to evaluate each kind of Nav system performance.

Table 1. Proposed questionnaire for the evaluation of the opinion of the drivers in using the three different navigation interfaces (Note All questions are translated from Italian language original)

1	Has it been easy to follow the route?
2	Were the indications clear?
3	Were they given early enough?
4	Were the turning points easy to recognize?
5	Is the terminology used easy to understand?
6	Had followed the street instead of the path on the screen?
7	Was the information enough?

2 Results and Main Considerations

According to the results of the first questionnaire all the subjects resulted well trained in the use of navigation systems (while driving) and of electronic interfaces (internet, cell phone, …). It is important in particular to highlight the results of the question n. 6. The drivers recognized that the 2-Me navigator interface allows them to follow the road more and consult the navigation screen less.

Concerning the number of the head movements done and the time spent to look to the navigator system, the 2-Me and C-Me systems produce a lower number of movement towards the navigator, i.e. the trip is better understood in these two modalities with respect to the Sat-Nav. The new interfaces (2-Me and C-Me) also reduced glance away and use of the navigation display information thanks to the introduction of landmarks. The literature shows that any single display glance longer then 2.5 sec is inherently dangerous [16]. The safest of the three tested navigation systems seems to be the **2-Me** interface. In this way we confirm that the tests presented in this paper resulted in an improvement of the navigation system interface.

3 Discussion NEV Application

Current navigation systems display a density of visual detail that do not match a driver's perception of the dynamic road environment from inside the moving vehicle cabin. This mono sensory input produces a disconnect as the driver tries to interpret an abstracted reality which has already changed in the glance away time from road to display and back again. This was demonstrated in our test results; drivers using standard navigation maps showing an increased propensity to look at the screen. What is already a problem in standard maps risks becoming a serious safety issue with data dense Electric Vehicle navigation systems. The DHS proposal is to restructure the map into a multimodal dynamic data feed which better matches the driver perceived external environment. The main characteristics of the new **2-Me** navigation interfaces consists of **context based commands.** Turn right at the "church" or at a "school" instead of "turn right after 140 meters" or other distance referenced commands. In this way landmarks, (lights, banks or commercial buildings, etc.) quickly identifiable in an urban landscape, in a driving context, are easily recognized by the driver. They are also supported by vocal commands. More advanced systems would “internalize” the landmarks by inserting personal references : “turn left at Toms house”. DHS would reinforce the multimodality (visual/voice) with tactile or force feed back prompts integrated with this data feed. The mass adoption of Electric Vehicles will bring significant environmental benefits, but also interaction and energy management challenges. This is even more the case with fleet operated Networked Electric Vehicles. These have multiple destinations and time sensitive (energy usage optimized) routes. As a result a careful structuring of navigation data needs to be done to improve usability in relation to safety and also address “range anxiety” concerns. DHS believes that the current “map view“ model is an outdated if not misapplied in car solution. We believe the **2-Me** concept offers an alternative modality for use in dynamic navigation environments. It is not a finished product and the testing has posed further questions on how the multimodal data feed can be further expanded and more finely calibrated to improve usability. It has also posed questions on how a dynamic and adaptive system can calibrate itself to gender and age variables. These items we wish to address in our next round of testing.

References

[1] Richtel, M., et al.: NY Times/2009/07/19/technology

[2] Ward, et al.: The Evolution of thought Chapter 18. Body size and intelligence in Hominoid evolution. Department of Anthropology, Department of Pathology and Anatomical Sciences, University of Missouri, Columbia, MO 65211, USA, pp. 335

[3] Bowden, J., et al.: The Most Effective Natural Cures on Earth: The Surprising, Unbiased Truth about What Treatments Work and Why

[4] Wittmann, M., et al.: The neural substrates of subjective time dilation Front. Hum. Neurosci. (2010)

[5] Zenati, N., et al.: ev. Centre of Adv. Technol., Baba Hassen, Algeria (2007) ISBN: 978-1-4244-1235-8

[6] Strayer, Drews, Crouch, et al.: A comparison of the Cell phone driver and the drunk driver (2006)

[7] Stutts, J.C., Knipling, R.R., Pfefer, R., Neuman, T.R., Slack, K.L., Hardy, K.K.: Guidance for implementation of the AASHTO strategic highway safety plan: A guide for reducing crashes involving drowsy and distracted drivers. Transportation Research Board, Washington, D.C. (2005)

[8] McKnight, A.J., Mcknigh, A.S.: The effect of cellular phones use upon driver attention. Accident Analysis and Prevention 25, 259-965 (1993); Strayer, D.L., Johnston, W.A.: Driven to distraction: dual –task studies of simulated driving and conversing on a cellular phone. Pysic Sci. 12, 462–466 (2001)

[9] Alm, H., Nilsson, L.: Changes in driver behaviour as a function of hands free mobile phones – a simulator study. Accident Analysis and Prevention 26, 441–451 (1994)

[10] Dingus, T.A., Hulse, M.C., McGehee, D.V., Manakkall, N., Fleischmann, R.N.: Driver performance results from the Trav –Tek IVHS camera car evaluation study. In: Proceeding of the Human Factors and Ergonomics Society 38th Annual Meeting, pp. 1118–1122. Human factors and Ergonoimcs society, Santa Monica (1994)

[11] Dingus, T.A., Hulse, M.C., Mollenhauer, M.A., Fleischmann, R.N., McGehee, D.V., Manakkall, N.: Effects of age, system experience, and navigation technique on driving with an advanced traveler information system. Human Factors 39, 177–199 (1997)

[12] Newell, et al.: Perception 36, 1415–1417 (2007), New insights into multisensory perception

[13] Horrey, W.J., Lesh, M.F., Garabet, A.: Assessing the awareness of performance decrements in distracted drivers. Accident analysis and Prevention 40, 675–682 (2008)

[14] Hancock, P.A., Lesch, M.F., Simmons, L.: The distraction effect of phone use during a crucial driving maneuver. Accident Analysis and Prevention 35, 501–514 (2003)

[15] Lees, M.N., Lee, J.D.: The influence of distraction and driving contex on dirver response to imperfect collision warning systems. Ergonomics 50, 1264–1286 (2007)

[16] Bhise, V.D., Forbess, L.M., Farber, E.I.: Driver behavioral data and considerations in evaluating, Washington D.C (1986)

[17] National Highway Traffic Safety Administration (NHTSA), http://www.suvblogger.com/car-safety-guide/are-you-a-distracted-driver/

Prospecting a New Physical Artifact of Interaction for iDTV: Results of Participatory Practices

Leonardo Cunha de Miranda, Heiko Horst Hornung, and M. Cecília C. Baranauskas

Institute of Computing, University of Campinas (UNICAMP)
13083-852, Campinas, SP, Brazil
professor@leonardocunha.com.br, heix@gmx.com,
cecilia@ic.unicamp.br

Abstract. A literature review has indicated that the remote control, the main physical artifact of interaction with the television system, in its current form is not adequate to the interaction between users and applications of Interactive Digital Television (iDTV), especially in a scenario of diverse user profiles as found in Brazil. This paper describes participatory practices carried out with the intention of defining a new physical artifact of interaction for iDTV. Based on the results of these participatory practices and previous research results, we provide a definition of an artifact that can be adapted to diverse contexts of use.

Keywords: interactive digital television, interaction design, participatory design, digital artifact, gesture-based interaction, human-computer interaction.

1 Introduction

We have been investigating the design of interactions of users with Interactive Digital Televison (iDTV) focusing on the physical artifacts of interaction since 2006 [13]. Starting from earlier results [10,11,12], this paper aims at presenting results of participatory practices that lead to the definition of a new digital artifact and its interaction language.

Gawlinski [5] defines interactive television as a set of technological devices that allow the establishment of a dialogue between the user or viewer with a TV channel, program or service. In its current form, the remote control, the main physical artifact of interaction with the television system, is not sufficient for a more continual and dynamic interaction of users with the iDTV, considering the problems already identified and discussed in literature by various authors, e.g. [2,3,4,8,14]. According to Rocha and Baranauskas [17], the same technology that simplifies life by aggregating a number of features into a single object, makes life more complicated by demanding efforts for learning and using technology. This is the paradox of technology and the challenge is to minimize these effects. Therefore, we emphasize the need to develop or adapt new physical artifacts of interaction for iDTV in order to take advantage of this new interactive media.

While Brazil has adopted one of the worldwide DTV Standards as a technological basis for Brazilian System of Digital Terrestrial Television (SBTVD-T), this does not mean that SBTVD-T meets all our needs in terms of how the user interacts with

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 167–176, 2011.

television. The context of use of digital television in Brazil is different from the developed countries, because we live in a context of large socioeconomic and cultural differences, also with regard to access to technology experience and knowledge in different regions of the country. Within this scenario, we believe it is essential to analyze and propose artifacts to facilitate user interaction with iDTV, and maximize the use of such new media as an important tool for making information and knowledge available and accessible to all citizens.

This paper is organized as follows: Section 2 presents the methodology used for this work; Section 3 describes the results of participatory practices; Section 4 discusses these activities; and Section 5 presents concluding remarks and provides pointers to future work.

2 Methodology

We understand that knowing the needs of users and their context of use is very important for proposing artifacts suitable for everyone. Employing techniques of Participatory Design (PD) [18], we hence promote the active participation of representatives of the target audience during the process of defining a new interaction language for a new physical artifact of interaction with iDTV. PD was used as the methodological foundation of practices whose goal was to design and explore the interaction language for a new digital artifact for iDTV in a participatory way with end-user representatives.

This approach involves user participation in all stages of development, not just during the test phases of prototypes or evaluation. In general, this approach has three important characteristics: i) it is oriented to the context of the target audience; ii) it involves collaboration; and iii) it is iterative. Thus, this approach permits the design of solutions <u>with</u> users, and <u>for</u> users. This approach has been used successfully in several studies, including scenarios of diversity of users, e.g., in [1,16].

From the set of participatory methods and techniques, we used the technique of brainstorming, since it is a simple technique that requires few resources. The brainstorming sessions were held at different times and conducted by the first author of this paper. In this work, these activities are referenced as Participatory Brainstorming Practices (PBP). As to the dynamics of the PBPs, participants sat on chairs arranged in a semicircle in front of a whiteboard, which allowed that ideas uttered and discussed by participants were written or drawn on the whiteboard by themselves.

The participatory practices were related to the formalization of a new gesture-based interaction language for iDTV considering a new physical artifact of interaction. Our interest was to find out a language that would represent some actions to interact with iDTV, e.g., turn the TV on/off, adjust the sound volume, change channels, and activate a particular interface element. In general, the goal of the participatory practices were to share and discuss potential solutions, without restricting to only those proposals, and asking users to show possible inconsistencies of the interaction language, e.g., by means of counter-examples.

It is worth mentioning that during the participatory practices we tried not to emphasize possible limitations of the technology in order to not restrict participants' ideas, e.g., regarding the question of how to provide enough electrical current to power a miniature radio frequency (RF) transmitter.

3 Participatory Practices

The participatory practices were conducted in Campinas (São Paulo, Brazil) in November/December 2008 and were attended by five users. We chose to initially perform these activities with a group of users with experiences in the use of digital artifacts and digital media in general, in order to subsequently present a refined proposal for discussion during participatory practices with other groups of users with greater diversity regarding skills and competencies, as well as access to technologies. This incremental approach enables the participation of different segments of the target audience during different stages of the design process, while keeping practices manageable and without needing to repeat the same practices with different participants.

During participatory practices, we explored artifacts that are not – yet – commonly used for interaction of users with digital systems in general. A point of departure for the participatory practices was prior knowledge about the artifacts identified during our analysis of interaction with iDTV presented in [11]. Consequently, during the participatory practices, mock-ups were used as a way to simulate the tasks that can be performed applying the interaction language with these artifacts. The artifacts used in the participatory practices are presented in Fig. 1.

(a) (b) (c)

Fig. 1. Artifacts used during participatory practices (a) pair of gloves used during the first practice (b) thimbles used during the second practice (c) rings used during the third practice

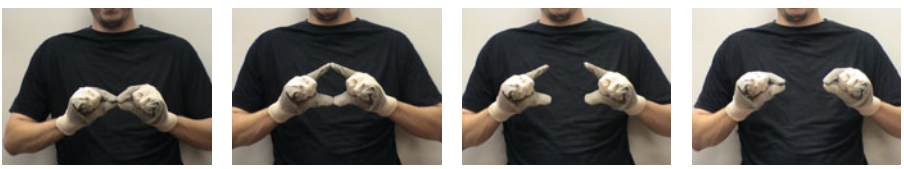
3.1 The First Practice

Our initial idea for the solution involved two components: an artifact for gesture-based interaction and a capturing device for reading the movements produced during interaction. Regarding the first component, we chose as a first approach to explore, the use of gloves in order to verify if this device would fit the context of use. Fig. 1a presents the gloves used during the first practice; the gloves would later have to be electronically adapted in order to work with the capturing device.

After defining the glove as physical artifact, the next step was to define the language of gesture-based interaction enabled by this artifact. Our initial strategy was to define a set of gestures that could represent the different actions needed for interaction with iDTV.

Regarding the conduction of the actual practice, the facilitator briefly explained the research context to the participants, and subsequently presented the interaction language in order to discuss with the participants whether the language could actually be used for interacting with iDTV. The initial idea was to define the gestures of the interaction language considering up to four points of the Multi-Touch Imaginary Screen (MulTIS) Model (for further details see [10]). Considering motion in 2 axes – X and Y – and disregarding the Z-axis, each glove has mapped two points, resulting in four points for two gloves. The definition of gestures should consider relative movements of these points without considering the temporal factor, which we understood to allow the composition of a new language for interaction with the iDTV through the glove as an artifact of physical interaction.

To exemplify, Fig. 2 shows a sequence of gestures for switching on the television using the gloves. The sequence has been suggested by the users participating in the first practice. The proposal was to join the four points, two in each hand (Fig. 2a), and perform the following motion: moving the upper points upward and lower points downward while holding the respective upper and lower points together (Fig. 2b); separating the respective upper and lower points horizontally to the left and right, marking four corners of a rectangle (Fig. 2c); finally, joining the respective left and right dots (Fig. 2d). The proposal to turn off the television was the reverse sequence of movements.



(a) (b) (c) (d)

Fig. 2. Sequence of gestures for switching on the TV using the gloves

At the end of this practice we concluded that a language of interaction based on four points could be complex, since, for example, the need for four cursors on the screen to map the four points could cause difficulties in understanding and using the gloves for users with motor impairments. Hence, using counter-examples, and considering the discussion of television use by the target audience, the participants of this first practice, demonstrated the difficulty of formulating a language of interaction based on four points. Furthermore, it became clear that the solution should consider fewer points for defining the gestures of the interaction language and perhaps consider the temporal dimension besides the spatial dimension. Up to the end of the first practice, the point of departure for defining the interaction language was the simplest functions, i.e., turning the TV on or off.

After completing the first practice, we realized how complex a formalization of a gesture-based language for the desired context would be, since movements during interaction can happen unintentionally. In addition, several issues with the gloves as physical artifact were raised, e.g. questions of hygiene, which led us to the conclusion that gloves did not seem to adapt well to the use context of iDTV. For the next practice we thus defined a different proposal for the interaction language and artifact.

3.2 The Second Practice

For the second practice, we kept the same idea of the solution architecture, i.e., dividing it into two devices, one being the physical artifact enabling the interaction between the user and the television system, and the other the capturing device for reading the gestures performed by the user.

Considering the discussion about the glove during the first practice, we proposed a new physical artifact of interaction for the second practice, that we understood to be more appropriate than the glove: the thimble. Fig. 1b shows the thimbles used during the second practice.

Initially, we envisaged the use of three interactive thimbles, or an electronically adapted version of the thimbles shown in Fig. 1b, whereas only two thimbles would be mapped to the iDTV interface. We assumed that this configuration would allow flexibility of use, because the users could choose the fingers they would wear the thimbles, as well as choosing to use the thimbles on one or on both hands. The idea of maintaining the mapping of four points was discarded due to the results of the discussions during the first practice and due to potential difficulties of users with severe motor restrictions.

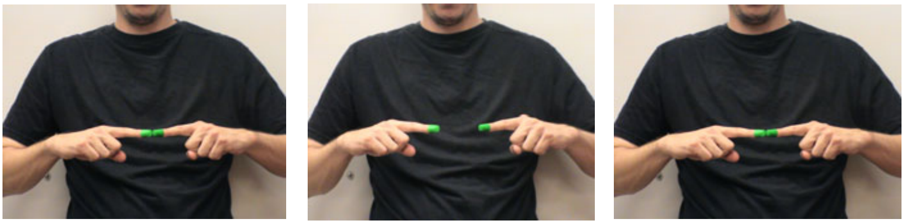
Although we employed gestures composed of two points during the second practice, we also took into account that the language should be usable if the two points were mapped to one hand instead of two, since within the context of television use, an interaction language that requires the use of both hands could have a low user acceptance.

Within this configuration, two of the three thimbles would be mapped as points of the MulTIS Model and used for the definition of the interaction language, which seems less complex than the four-point approach of the first practice, both in terms of use and implementation.

During the second practice, the syntax of the interaction language, that allows to represent the possible interactions with the iDTV, is composed by the relative movements of the two points and the temporal component. Taking into account a suggestion of one of the participants of the first practice, besides the spatial dimensions, i.e. rotations, approximations, this approach also considers the temporal dimension. Furthermore, at this stage of our research we also explored if real-life metaphors could inform the definition of a gesture-based interaction language.

To exemplify, Fig. 3 shows the sequence of gestures to turn on or off the TV using the thimbles, which was suggested by the users participating in the second practice. The proposal was to join two points and maintain them joined for three seconds (Fig. 3a), then to horizontally separate these points keeping them horizontally aligned for more than three seconds (Fig. 3b). After that, the two points are rejoined at the center (Fig. 3c). One problem identified in this proposal is that it would take nine seconds to turn on or off the television while when using a current conventional remote control, the response would be almost instantaneous.

At the end of the second practice, the point of departure for defining the interaction language still was exploring the simplest functions regarding interaction, i.e., turning the TV on or off. While exploring simple functions during the practices, the authors already started to explore more complex functions, such as manipulation of user



(a) (b) (c)

Fig. 3. Sequence of gestures for switching on/off the TV using the thimbles

interface elements. Possible gestures for those more complex functions often conflicted with simpler functions explored during the practices. Hence, at this point we realized that it was necessary to already consider those more complex functions.

As an alternate artifact, the use of band-aids was suggested during this practice. However, it was discarded for representing convalescence, which demonstrates the need for a more accurate understanding of social signifiers [15] regarding the artifacts. Nevertheless, the band-aid idea influenced the choice of the artifact to be explored in the third practice.

We observed that during this practice, issues were more related to the physical artifact of interaction and much less on the language of interaction. However, from our observations we concluded that the temporal component would bring more problems than benefits, so it should be used with caution.

3.3 The Third Practice

For the third practice we continued to use the separation between artifact of interaction and capturing device. However, as a new artifact of interaction for this practice we introduced the ring which should be adapted to electronically identify mechanical pressure, i.e. if the ring was pressed. The initial proposal did not contemplate the inclusion of any button on the device. Fig. 1c shows the rings used during the third practice.

The solution of the ring with a button, which we call interactive ring, responded to the problem we observed during the previous two practices, that the command in the MulTIS Model to activate a user interface element would be represented by the gesture of joining two points. This initial definition was unfeasible because of conflicts with other commands. Moreover, the solution of the rings also presented some advantages over the thimble, where the user's finger had to be pointed at the television screen.

One of the motivations to include a button was a comment of a participant of the second practice: "I don't see any movement that could represent the action of turning off the TV without pushing an on-screen button like in Windows and Linux". At this point, the facilitator of the participatory practices noted that creative solutions, for example the suggestion made by a user of the second practice to turn on/off the TV with a rotation of the points, show that creative ideas can indeed provide solutions to problems that seem to be hard to solve.

Aiming to design solutions for all functions of interaction with iDTV, and considering the problems identified during the two previous practices in the initial proposal of the third practice was to divide the rings into three distinct types: i) for movement; ii) for activation; and iii) for options. The three types of rings would be physically adjustable, colored, with identification in Braille[1] and a simple and easy to learn interaction language.

To exemplify, Fig. 4 shows how to turn on or off the TV using the rings, i.e. to turn on or off the television, the button of the options ring – blue – has to be pressed for at least two seconds.

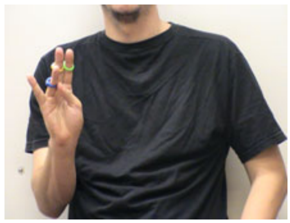


Fig. 4. Turning on/off the television using the rings

At the end of the third practice we found that an interaction language for iDTV based exclusively on gestures would be complex in order to avoid conflicts in commands. In this environment a Command Mode[2] would probably be necessary. However, with the advancement of the physical artifact of interaction, i.e. the introduction of a button, the proposal resulted in a simple gesture-based interaction language using rings with only one button, and without the need of a special Command Mode. During the third practice a bracelet instead of the rings was suggested too. However, during the course of the practice, we perceived that the physical location of the bracelet on the wrist would not allow as much flexibility of interaction. For example, the need of having a button on the artifact, could cause problems during interaction.

4 Discussion

After concluding the three participatory practices, it became clear that, due to numerous factors that influence the television use, it is difficult to find a new

[1] The Braille stamp was conceived as a means of facilitating the identification of the rings for visually impaired users.

[2] The Command Mode was defined during the second practice to minimize problems with unintentional gestures made, e.g., by users during a conversation in front of the TV. In order to be able to distinguish between intentional gestures that should trigger a command, and unintentional ones, one proposed solution was to introduce a gesture to enable or disable the so-called Command Mode. Upon activating this mode, the gestures would be mapped by the system as gestures of interaction upon disabling this mode, the system would no longer consider the gestures made by users.

interaction language for the studied context. This same consideration is valid regarding proposals for gestures as an additional mechanism for data entry in iDTV.

During the first practice, the little elaborated physical artifact of interaction did not allow the definition of an interaction language without conflicts. The complexity induced by the four-point approach of the physical artifact made it difficult to create a gesture-based language accessible to our target audience. Thus, during the second practice, we noticed that the emphasis of the work was directed almost exclusively to the formulation of gestures that could constitute the interaction language, in order to define the movements that would represent the different functions of interaction with the iDTV.

Between the first and the second practice, significant changes to the interaction language were made. However, the physical artifact of interaction did not evolve in the same order of magnitude, although different physical artifacts of interaction were proposed during each participatory practice. Upon finishing the second practice, we observed that the two aspects of this research – the physical artifacts of interaction and the gesture-based interaction language – were not explored in a balanced way. Thus, during the discussions held between the second and third practices, this new perspective was applied, which made a considerable difference in the outcomes of this work.

Table 1 presents the different artifacts explored during the participatory practices. As described in this table, each participatory practice has different settings for the artifacts, hands, and points mapped.

Table 1. Proposals for physical artifacts of interaction

PBP	Artifact	Hand	Points to be mapped
1st	2 gloves	2 hands	4 points
1st	2 gloves	2 hands	2 points
2nd	4 thimbles	2 hands	4 points
2nd	3 thimbles	2 hands	3 points
2nd	2 thimbles	1 hand or 2 hands	2 points
2nd	2 band-aids	1 hand or 2 hands	2 points
3rd	3 rings	1 hand or 2 hands	1 point
3rd	1 bracelet	1 hand	1 point

In selecting the rings as physical artifact of interaction we are not invalidating the use of other artifacts – glove, thimble, band-aid, and bracelet – as potential instruments of interaction of users with digital systems. However, considering the presentations and discussions raised during the participatory practices, these artifacts revealed not the most appropriate ones, regarding one of our research goals, namely the specification of a new physical artifact and an interaction language that can be used by all, to the greatest extent possible. The ring for example has shown to be more ergonomic in our context than the other artifacts.

Participatory practices contributed significantly to reflect on ideas and to a better understanding of the factors that influence the context of use. The practices have enabled us to better explore our proposals and to design and evaluate them together with representatives of the target audience. We also understand that the evolution of the solution gained another momentum with the active participation of users in the initial phase of defining a new interaction language. Furthermore, this approach allowed a better understanding of the daily practices of the target audience, which facilitated the search of the artifact and the definition of the language, and enabled innovation inspired by the diversity present in the Brazilian population. Detailed results of the practices are available in [7].

5 Conclusion

Brazilian iDTV potential users are not yet familiar with the everyday use of digital interfaces. Activities carried out by our research group together with representatives of the target audience in other research contexts, demonstrated the difficulty of understanding the users of digital interfaces as well as the difficulties users have understanding digital interfaces. Paths to effective use and a more fluent dialogue with this new medium will depend directly on the physical artifact of interaction with iDTV.

This work presented participatory practices conducted with users to identify an artifact for the interaction of users with iDTV applications. Based on what was discussed during these activities, we chose to continue the work with the proposal of the ring. Thus, the goal originally proposed for this work was reached when we defined an artifact and clarified the iDTV use context with its implications on the design of future solutions.

We believe that a more direct interaction with iDTV, requires the physical artifact of interaction to be more transparent to the user. Thus, the focus of interaction must be directed to the user interface of the interactive applications and not to the interaction artifact itself. A gesture-based interaction language based on the MulTIS Model via rings provides a more natural interaction in accordance with the interaction desired by users.

As a continuation of this research are already completed the specification of the product design and the interaction language of a new digital artifact for iDTV based on the concept of a ring. We further elaborate on the formalization of the interaction language; detailed description of the physical artifact and additional resources are presented elsewhere [6,9].

Acknowledgments. This research was partially funded by the National Council for Scientific and Technological Development – CNPq (grants #141489/2008-1 and #141058/2010-2), by the Brazilian Federal Agency for Support and Evaluation of Graduate Education – CAPES (grants #01-P-09881/2007 and #01-P-08503/2008), and by the State of São Paulo Research Foundation – FAPESP (grant #2010/11004-9).

References

1. Almeida, L.D.A., Neris, V.P.A., Miranda, L.C., Hayashi, E.C.S., Baranauskas, M.C.C.: Designing Inclusive Social Networks: A Participatory Approach. In: Ozok, A.A., Zaphiris, P. (eds.) OCSC 2009. LNCS, vol. 5621, pp. 653–662. Springer, Heidelberg (2009)

2. Berglund, A.: Augmenting the Remote Control: Studies in Complex Information Navigation for Digital TV. Doctoral Thesis, Linköping University (2004)
3. Cesar, P., Bulterman, D.C.A., Jansen, A.J.: Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content. In: Tscheligi, M., Obrist, M., Lugmayr, A. (eds.) EuroITV 2008. LNCS, vol. 5066, pp. 168–177. Springer, Heidelberg (2008)
4. Cesar, P., Chorianopoulos, K., Jensen, J.F.: Social Television and User Interaction. ACM Computers in Entertainment 6(1), 1–10 (2008)
5. Gawlinski, M.: Interactive Television Production. Focal Press, Oxford (2003)
6. Miranda, L.C., Baranauskas, M.C.C.: Anéis Interativos Ajustáveis: Uma Proposta de Artefato Físico de Interação para a TVDI. In: 4th Latin American Conference on Human-Computer Interaction, pp. 16–23. UABC, Ensenada (2009)
7. Miranda, L.C., Baranauskas, M.C.C.: Prospectando um Novo Artefato Físico de Interação para a TVDI. Technical Report, University of Campinas (2010)
8. Miranda, L.C., Hayashi, E.C.S., Baranauskas, M.C.C.: Identifying Interaction Barriers in the Use of Remote Controls. In: 7th Latin American Web Congress, pp. 97–104. IEEE Computer Society, Los Alamitos (2009)
9. Miranda, L.C., Hornung, H.H., Baranauskas, M.C.C.: Adjustable Interactive Rings for iDTV. IEEE Transactions on Consumer Electronics 56(3), 1988–1996 (2010)
10. Miranda, L.C., Hornung, H.H., Baranauskas, M.C.C.: Prospecting a Gesture Based Interaction Model for iDTV. In: IADIS International Conference on Interfaces and Human Computer Interaction, pp. 19–26. IADIS Press, Lisbon (2009)
11. Miranda, L.C., Piccolo, L.S.G., Baranauskas, M.C.C.: Artefatos Físicos de Interação com a TVDI: Desafios e Diretrizes para o Cenário Brasileiro. In: VIII Brazilian Symposium on Human Factors in Computing Systems, pp. 60–69. SBC, Porto Alegre (2008)
12. Miranda, L.C., Piccolo, L.S.G., Baranauskas, M.C.C.: Uma Proposta de Taxonomia e Recomendação de Utilização de Artefatos Físicos de Interação com a TVDI. In: Workshop on Perspectives, Challenges and Opportunities for Human-Computer Interaction in Latin America, pp. 1–14 (2007)
13. Miranda, L.C.: Artefatos e Linguagens de Interação com Sistemas Digitais Contemporâneos: Os Anéis Interativos Ajustáveis para a Televisão Digital Interativa. Doctoral Thesis, University of Campinas (2010)
14. Nielsen, J.: Remote Control Anarchy, `http://www.useit.com/alertbox/20040607.html`
15. Norman, D.A.: Signifiers, Not Affordances. ACM Interactions 15(6), 18–19 (2008)
16. Okamoto, M.: Possibility of Participatory Design. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 888–893. Springer, Heidelberg (2009)
17. Rocha, H.V., Baranauskas, M.C.C.: Design e Avaliação de Interfaces Humano-Computador. NIED/UNICAMP, Campinas (2003)
18. Schuler, D., Namioka, A.: Participatory Design: Principles and Practices. Lawrence Erlbaum Associates, Inc., Hillsdale (1993)

Optimisation of Sound Localisation for Emergency Vehicle Sirens through a Prototype Audio System

David Moore, Stephen Boslem, and Vassilis Charissis

Glasgow Caledonian University, School of Engineering and Computing,
Department of Computing and Creative Technologies, Cowcaddens Road, Glasgow, UK
J.D.Moore@gcu.ac.uk, sbosle10@caledonian.ac.uk,
v.charissis@gmail.com

Abstract. This paper examines the issues associated with the localisation of emergency vehicles. A combinatory warning system is then proposed that aims to provide drivers of both civilian and emergency vehicles with a different sequence of auditory cues as well as an in-cabin warning when an emergency vehicle is in the close vicinity. For the early testing of this hybrid alert system, we used the modelling techniques currently available to the UK emergency services in order to estimate the concurrent efficiency of the siren's auditory warnings.

Keywords: Road Safety, Sound Localisation, Warning Systems, Ambisonics, Spatial Audio.

1 Introduction

Emergency vehicle (EV) warning systems currently consist of three forms of alert: an audible warning, flashing lights and coloured bodywork markings. Out of the three, the most prominent system is the audible warning, which is typically based on a sweeping siren pattern. Early localisation of sirens by members of the public makes safer and swifter manoeuvring through traffic possible. However, for pedestrians and motorists alike, sirens can cause confusion, disorientation and possible danger if not reacted to within a timely manner. According to the Accident Statistics published by the British Department of Transport Road, the driver's inability to locate the incoming EV resulted in 7 fatalities and 1,226 casualties in the UK in 2008 [1].

Today, cars have improved soundproofing and are often equipped with powerful stereos. These factors combined with the general cacophony of modern urban life are enough to mask alert sounds, making siren recognition and accurate localisation challenging. Previous work has shown that high levels of background noise can significantly affect the average driver's ability to clearly define the position of the EV [2]. It has also been shown that the human inability to accurately locate the direction of an approaching EV is linked to the siren patterns and the limitations of the human auditory system [3].

As roads become increasingly busier it is important to design and implement systems that alert the driver of the proximity of incoming emergency vehicles and, more importantly, which direction they are approaching from. In this work we

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 177–186, 2011.

investigate the issues of localising EVs and in turn propose a new combinatory approach that provides the drivers with a different sequence of auditory cues as well as an in-cabin alert presented when an EV is in the close vicinity.

The section following looks at research into sound localisation as well as current siren localisation issues. Then the "Proposed System" section offers a succinct overview of the prototype combinatory approach. Then follows information on the modelling and simulation scenarios used to evaluate the effectiveness of the proposed system in contrast to the existing setup. The results are presented and then the conclusion section outlines the limitations and considerations of the proposed system and presents a plan outlining how we intend to develop our system in future work.

2 Siren Localisation Issues

2.1 Sound Localisation

Sound can be the first warning of events that are taking place around us. Unlike vision, which can be obscured by buildings and other objects, sound is perceptible from all directions and can travel around objects to some extent. Thus, auditory information is vital for situation awareness on the roads, both as a complement and a supplement to vision.

The human auditory system uses both binaural cues and monaural cues for determining the angle and distance of a sound source in the horizontal and vertical planes [4]. Binaural cues occur because we have two ears separated by the width of our head. Hence, any sound that originates from either side of our head will arrive at the ear closest to it before reaching the other ear; it will also be louder at the closest ear. These cues are known as Interaural Time Difference (ITD) and Interaural Level difference (ILD) respectively. ITDs are not an effective cue at high frequencies because certain time delays may result in the same phase difference at the ears. Conversely, ILDs are not effective at low frequencies because a sound's wavelength is larger than the diameter of the head resulting in little or no attenuation by the head when en route to the furthest ear.

Although ITD and ILD are known to be the major cues for sound localisation they are not enough on their own for localising sounds. For sound sources in the horizontal plane there are always two points around the listener with identical ITDs and ILDs. For example, sound arriving from a source at 45° from the front in the horizontal plane will have an identical ITD and ILD as sound arriving from a source at 135° from the front. If the vertical plane is considered as well then there will be a whole series of points on the surface of a cone that have the same ITD and ILD. This is known as the cone of confusion (see Fig.1).

To resolve this ambiguity spectral cues are used which occur as a result of the directional-dependent filtering caused by sound reflecting off the ear's pinnae and upper body. A number of different studies have demonstrated that monaural (single ear) spectral cues are vital for the localisation of sound sources above and below the listener [5, 6]. This has been clearly demonstrated in experiments by Gardner and Gardner [7]. There is also substantial evidence that the spectral cues incurred because of sound reflecting off the pinnae help us discriminate sounds coming from the front and back [8-10].

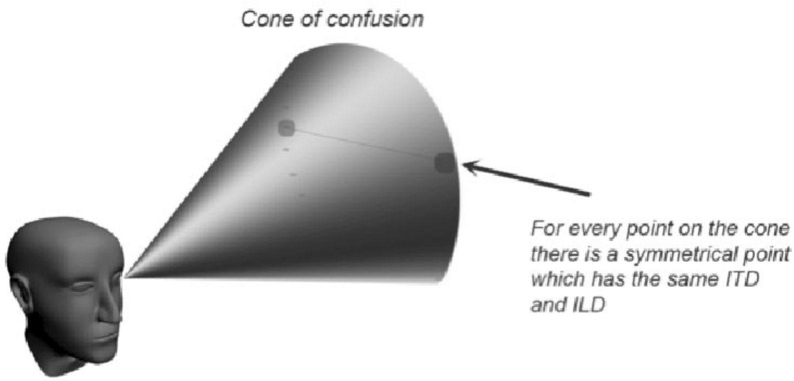


Fig. 1. The cone of confusion

Numerous studies have shown that human localisation accuracy varies markedly with frequency [11, 12]. Generally, human localisation accuracy remains approximately constant for frequencies below 1 kHz. For frequencies between about 1 kHz and 3 kHz, however, acuity degrades somewhat until after 3 kHz when it improves again. The reason for degradation in this frequency region is because interaural phase cues start to become ambiguous after 1 kHz, whereas below 3 kHz the interaural level differences are not always significant enough for a listener to lateralise a sound successfully. This problematic cross-over region can be seen in various studies [4, 12].

Localisation has been shown to be most accurate directly in front of the listener [4]. This accuracy decreases as the source moves to the side of the listener and improves again at the direct rear. The relationship between the angle of the sound source and accuracy of localisation is approximately the same for both low and high frequencies. When the sound source is moving (as in the case of an approaching EV) this can add greater complexity to the task of sound localisation. Research has shown that humans can be quite inaccurate when determining the absolute location of a moving sound at a particular time [13, 14].

2.2 Existing Siren Issues

All of the above physiological issues have a direct impact on the driver's interaction with emergency vehicles. A siren can only be accurately localised if it contains sufficient localisation cues and these cues are audible in the environment in which the source is produced. Existing sirens typically employ a sweeping pattern with the majority of sound energy in the region of 500 Hz to 1500 Hz - this can clearly be seen in the time-frequency plot provided below (see Fig. 2). However, this pattern is not necessarily the easiest source to localise. Sounds of greater than 500 ms duration, with broadband spectral content and strong onsets are known to be easier to localise accurately than narrowband or tonal sounds with gradual onsets [15].

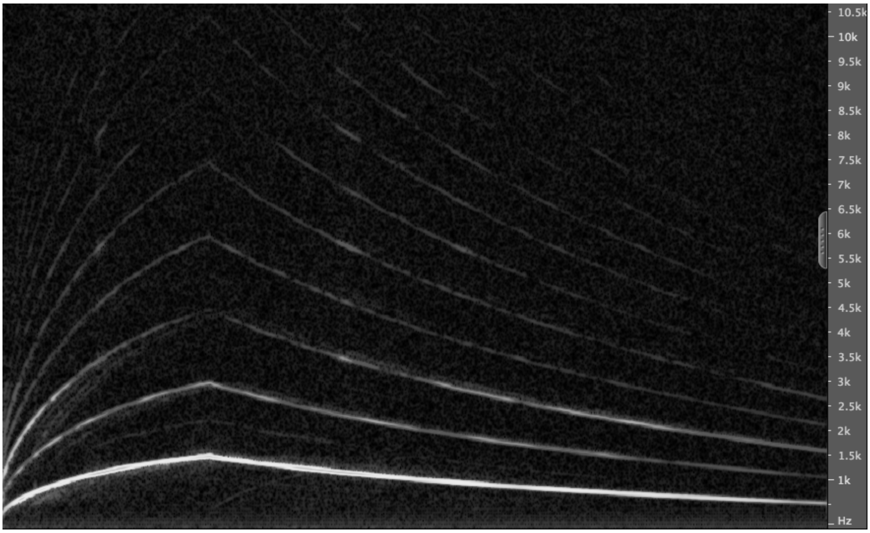


Fig. 2. Typical sweeping siren pattern

A previous investigation into audible warning systems showed that sirens were used 86% of the time by EVs when on call, but this was not always in unison with warning lights [16]. In a survey that examined public reactions to audible warning the results showed that nearly one third of drivers failed to detect the approaching emergency vehicle until it was less than 50m away. Even more astoundingly, 25% of the participants were completely unable to hear the audible warning systems on approach of the emergency vehicle. These figures again reinforce the limited effectiveness of current emergency vehicle audible warning systems [16, 17].

The aforementioned observations have clearly stated that some drivers could not differentiate the siren sound from the ambient noise at these distances, and evidently in some cases existing sirens are unable to overcome the background noise generated by driving on modern road networks, rendering them completely unusable in some cases [17-19]. Of all interactions with emergency vehicles, more than half of participants had to manoeuvre to let the vehicle pass. This discovery led to the so-called 'wake effect', where one third of these manoeuvres placed a third party at risk [20, 21].

3 Proposed System

The first improvement aims to minimise or even eliminate spatial ambiguity inherent in existing frequency siren tones. Therefore one of the siren patterns that we use in this study is designed to produce short, wide-band noise bursts that are known to be easier sources to localise than single frequency sweeps. Earlier investigations suggested that signals that possess relatively few harmonics are less resistant to masking, and they also need to be presented at much higher signal-to-noise ratio than signals that are acoustically richer [22, 23].

Adhering to this observation, siren patterns should contain the broadest audible frequency range possible in a range between 20 Hz to 20 kHz. Notably the main sweep of existing siren patterns only covers roughly the region between 500 Hz to 1500 Hz, therefore it is significantly more difficult for both motorists and pedestrians to accurately localise the source and direction of the sound [3, 17]. However the human hearing has a peak in sensitivity which reaches approximately 3500 Hz to 4000 Hz. To take account of the ear's increased sensitivity to this particular frequency band our proposed system uses a broader frequency bandwidth with a main sweep ranging from 400 Hz to 4000 Hz. It is proposed that by expanding the frequency content of existing patterns into the region of 3500 to 4000 Hz will make them more obvious, yet still sound familiar to motorists and pedestrians, building on the existing learned association of these tones.

The second element of the proposed solution aims to reduce the problems associated with environmental noise and the secluded shell of the vehicle interior. It consists of an in-cabin auditory and visual warning system for informing the driver when an emergency vehicle is in the vicinity. Nowadays, nearly all car audio reproduction systems consist of at least 4 loudspeakers arranged in a rectangle shape (see Fig. 3). This configuration is well suited to presenting auditory information at different spatial positions around the driver.

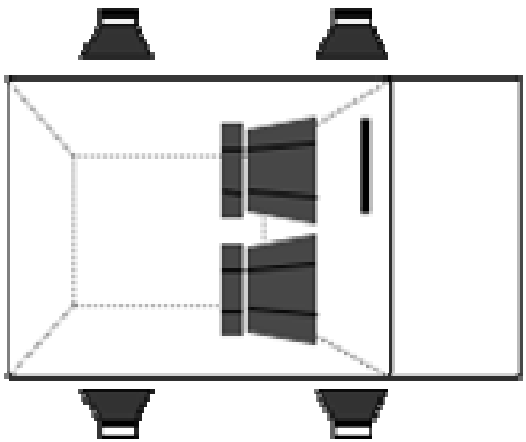


Fig. 3. Typical car loudspeaker layout

Therefore, we propose to relay a warning sound to the driver in the same direction as the approaching EV in order to reinforce its position. To do this, a surround sound technique known as Ambisonics will be employed to position the alert sound in the desired location. Ambisonics is currently an active topic of audio research and could be easily integrated with existing car audio systems. As the driver is not normally equidistant from the loudspeakers we intend to incorporate previous research by one of the authors that optimises the auditory localisation performance of Ambisonic surround sound for off-centre listening positions [24]. This second addition to the proposed system has been deemed essential for drivers that cannot hear and locate even the improved siren as has been highlighted. Several methods for automatically locating the EV before relaying a warning will be considered - see for example [25 - 27].

This two-part system aims to alert the driver simultaneously externally and internally to the EV. The leading vehicles in the traffic stream will then have a higher probability of identifying the alert and manoeuvre accordingly in order to support the safe passage of the EV. The preliminary experiments in the simulated environment demonstrated that a 30% speed increase could potentially be achieved with the combinatory proposed system. The predicted impact of increased road speeds on fatality rates and turn out times was examined using a simulation modelling method based on the FSEC modelling system. This is further discussed in the following section of the paper.

4 Modeling and Simulation

For the evaluation of the proposed solution it was deemed necessary to reproduce the effect of the system in a simulated environment and in turn contrast it to the impact of the typical auditory warning produced by the EVs. In this particular case, the Fire and Rescue Services of Scotland kindly provided the required information of current siren system. Furthermore we employed the official computer modelling methodology, which is used by the Central Scotland Fire and Rescue Services Fire Service Emergency Cover (FSEC), in order to achieve the validity of our comparative study.

This simulation, offered the comfort of experimenting in a safe, controlled environment. The simulation entailed different simulation scenarios typically modelled following real-life emergency situations which require the acute mobilisation of the Fire-brigade vehicular units. Notably the degree of realism embedded in these scenarios and factors that have been taken into account were crucial for the successful replication of reality in the simulation software.

In particular the base-cases, adhering to the standard speeds on the road network and consequently offer a truthful representation of the average road speeds typically attained by EVs with the use of the contemporary sirens. The main measurable data in this experiment was the distance that a fire-brigade vehicle can cover in a time defined array of 8 minutes in the selected region with centre the town of Falkirk. In turn the software measured the average speed achieved by the vehicles. The aforementioned two collected type of data could in turn define precisely the distribution of map coverage with and without the utilisation of the prototype system. Notably in the modelling process of the proposed system we included both the effect of the broadband noise patterns and RDS warning system in the simulation scenarios. This was feasible by utilising an existing trait of the FSEC software that allows the modification of individual road speeds, based on additional information.

As such we achieved to simulate the end-result of our system by predicting the increase in average speed in the specific road network. In turn both models were simulated and tested in the predefined scenarios with the use of the FSEC. The produced results were in turn visualised and propagated on the map, offering a visible differentiation on the EVs coverage with the use of both systems as shown in Figure 4 which illustrates the Ordnance Survey (OS) mapping in four distinctive map distributions varying from the standard road speed up to 30% increased speeds.

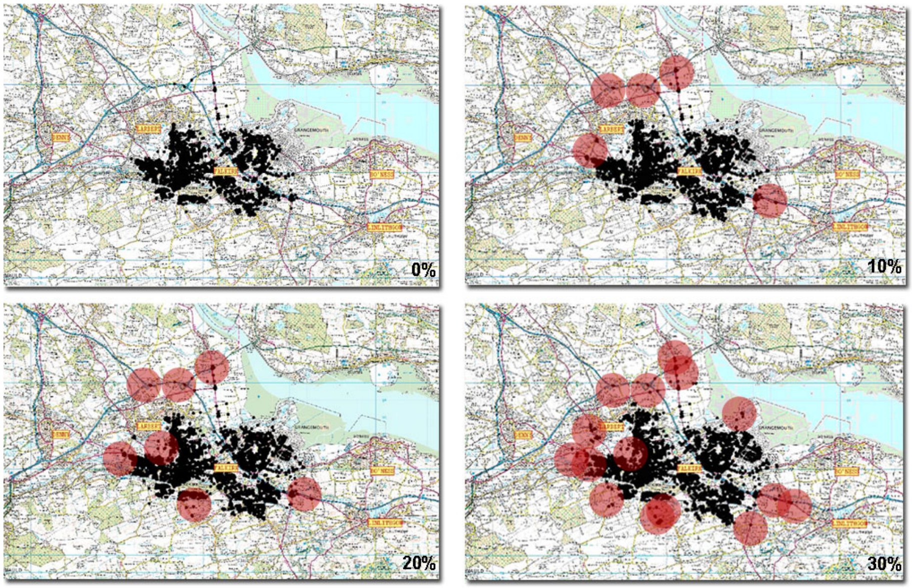


Fig. 4. Falkirk Station distance coverage in 8 minutes, under standard road speeds, 10% , 20% and 30% increased speed respectively

5 Discussion

It is evident that in the infinitesimal time-limit of 8 minutes, the EVs can extend their action significantly with the use of the proposed system, covering remote areas which previously were impossible to reach urgently.

However it could be argued that this method could be in expense of the civilians' safety as the EVs accelerate significantly more in order to achieve these distances in the same timeframe. This potential hazard can be decreased considerably as the proposed system does not inform the driver only through the external auditory avenue (i.e. siren), but presents the information also inside the vehicle with the use of the RDS. To this point the RDS can offer also visual cues provided in the instrumentation panel of the vehicle or through a Head-up Display (HUD) system. The latter could convey the information in driver's field of view, attracting instantly the user's attention. This could be utilised as a supplementary warning method in case of auditory cues' inability to command user's awareness.

Except the quantitative evaluation through the simulation that offers a clear view of the potential outcomes we also performed a qualitative study aiming to reveal the EVs' drivers feedback for the proposed system which will not be analysed in this paper. Nonetheless an important issue was raised as a direct result of using questionnaires to gather the opinions and experiences of this specific group of user, regarding the subject of driver behaviour towards emergency vehicle situations. Interestingly the emergency vehicle drivers stated that even if audible warning systems were substantially improved, "...some motorists will still panic and not know what to do, whereas others will simply not want to give way." As such a small segment of the driving population will still fail to handle the situation in a safe fashion even if an improved audible warning system can alert motorists to approaching emergency vehicles at an earlier stage than previously possible. The specific issue can be potentially resolved or minimised with the use of a combinatory approach of warning interfaces that can increase gradually, yet substantially, the alert levels (i.e. auditory, visual and tactile) depending on the distance difference from the EV and the approach speed.

6 Conclusions

This paper offered a succinct description of a prototype audio system which optimises effectively the sound localisation of emergency vehicles. The system entails an experimental sound sequence for the siren used in conjunction to an RDS warning signal. The proposed system aims to increase drivers' awareness regarding the incoming emergency vehicle and provide sufficient sound localisation that will enable the driver to respond promptly. In turn the successful provision of road space to the emergency vehicles will result in larger area coverage and faster service from the acute mobilisation units. The evaluation of the proposed system in contrast to the typical siren systems used in fire-brigade vehicles presented a potential benefit of 30% in the increase of the EVs speeds. This amplification of vehicular speed has a direct effect in a significantly expanded area that the emergency services can reach in the typical 8 minute window. The quantitative methods used for the evaluation were based on the modelling and simulation software used officially by the emergency services in Scotland.

On our future work we endeavour to test exhaustively the proposed system in a number of other simulation scenarios which will factor in different weather conditions, traffic flows and simultaneous accidents that require a predictive model of optimised navigation system. Furthermore we intent to embed the auditory cues of the proposed system on our existing full-windshield Head-Up-Display interface in order to provide the civilian drivers with a number of possible warnings depending on the level of emergency in each case.

References

1. Department for Transport UK, `http://www.dft.gov.uk/pgr/statistics/datatablespublications/accidents/`
2. Withington, D.J., Patterson, S.E.: Safer Sirens. Fire Engineers Journal 48, 6–10 (1998)
3. Withington, D.J.: The quest for better Ambulance Sirens. Ambulance UK 11, 20–21 (1996)
4. Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge (2001)
5. Wright, D., Hebrank, J.H., Wilson, B.: Pinna Reflections as Cues for Localization. The Journal of the Acoustical Society of America 56(3), 957–962 (1974)

6. Moore, B., Oldfield, S.R., Dooley, G.J.: Detection and Discrimination of Spectral Peaks and Notches at 1 kHz and 8 kHz. The Journal of the Acoustical Society of America 85(2), 820–836 (1989)
7. Gardner, M.B., Gardner, R.S.: Problem of Localization in the Median Plane: Effect of Pinnae Cavity Occlusion. The Journal of the Acoustical Society of America 53(2), 400–408 (1973)
8. Kistler, D.J., Wightman, F.L.: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. The Journal of the Acoustical Society of America 91(3), 1637–1647 (1992)
9. Langendijk, E.H.A., Bronkhorst, A.W.: Contribution of spectral cues to human sound localization. The Journal of the Acoustical Society of America 112(4), 1583–1596 (2002)
10. Zahorik, P.: Perceptual Recalibration in Human Sound Localization: Learning to Remediate Front-Back Reversals. The Journal of the Acoustical Society of America 12(1), 343–359 (2006)
11. Stevens, S.S., Newman, E.B.: The Localization of Actual Sources of Sound. The American Journal of Psychology 48(2), 297–306 (1936)
12. Moore, B.: An Introduction to the Psychology of Hearing. Academic Press, London (2003)
13. Perrott, D.R., Musicant, A.D.: Minimum Auditory Movement Angle - Binaural Localization of Moving Sound Sources. Journal of the Acoustical Society of America 62, 1463–1466 (1977)
14. Grantham, D.W.: Detection and Discrimination of Simulated Motion of Auditory Targets in the Horizontal Plane. Journal of the Acoustical Society of America 79, 1939–1949 (1986)
15. Vliegen, J., Optstal, A.J.V.: The Influence of Duration and Level on Human Sound Localization. Journal of the Acoustical Society of America 115, 1705–1713 (2004)
16. Saunders, G., Gough, A.: Emergency Ambulances on the Public Highway Linked with Inconvenience and Potential Danger to Road Users. Emergency Medical Journal 20, 277–280 (2003)
17. Ho, C., Spence, C.: Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of experimental psychology: Applied 11, 157–174 (2005)
18. Lenné, M.G., Triggs, T.J., Mulvihill, C.M., Regan, M.A., Corben, B.F.: Detection of emergency vehicles: Driver response to advance warning in a driving simulator. Human Factors 50, 135–144 (2008)
19. Llaneras, R.E., Green, C.A., Kiefer, R.J., Chundrlik, W.J., Atlan, O.D., Singer, J.P.: Design and evaluation of a prototype rear obstacle detection and driver warning system. Human Factors 47, 199–215 (2005)
20. Xiao, D.-l., Tian, Y.-j.: Reliability of Emergency Rescue System on Highway. In: Proceedings of Second International Conference on Intelligent Computation Technology and Automation (ICICTA 2009), Changsha, Hunan, China, vol. 2, pp. 564–567 (2009)
21. Sherratt, R.S.: Results of a DSP based adaptive cancellation filter system for the chirp, pulsar and localiser sirens. In: IEEE Nordic Signal Processing Symposium (NORSIG 2000), KolmArden, pp. 235–238 (June 2000)
22. Wightman, F.L., Kistler, D.J.: Sound localisation. In: Human Psychophysics, pp. 155–192. Springer, New York (1993)
23. Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L.: Localisation using non-individualised head-related transfer functions. Journal of the Acoustic Society of America 94, 111–123 (1993)

24. Moore, J.D., Wakefield, J.P.: Optimisation of the Localisation Performance of Irregular Ambisonic Decoders for Multiple Off-Centre Listeners. Presented at the 128th Audio Engineering Society Convention, London, UK (2010)
25. Valin, J.M., Michaud, F., Rouat, J.: Robust localisation and tracking of simultaneous moving sound sources using beam forming and particle filtering. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP (2006)
26. Atmoko, H., Tan, D.C., Tian, G.Y., Fazenda, B.: Accurate Sound Source Localization in a Reverberant Environment using Multiple Acoustic Sensors. Journal of Measurement and Technology 19(2) (2008)
27. Fazenda, B., Atmoko, H., Gu, F., Guan, L., Ball, A.: Acoustic Based Safety Emergency Vehicle Detection for Intelligent Transport Systems. In: Proceedings of the ICROS-SICE International Joint Conference Fukuoka, Japan, pp. 4250–4255 (2009)
28. Charissis, V., Papanastasiou, S., Vlachos, G.: Interface Development for Early Notification Warning System: Full Windshield Head-Up Display Case Study. In: Jacko, J.A. (ed.) HCII 2009. LNCS, vol. 5613, pp. 683–692. Springer, Heidelberg (2009) ISBN 978-3-642-02582-2

Applying Gestural Interfaces to Command-and-Control*

Todd Reily and Martina Balestra

The MITRE Corporation,
202 Burlington Rd 01730, Bedford / MA, United States
{treily,mbalestra}@mitre.org

Abstract. This whitepaper examines the applicability of gesture-based user interfaces in notional Command-and-Control (C2) environments of the United States Army. It was authored by a team of Human Factors Engineers at The MITRE Corporation, a not-for-profit research and development organization funded by the United States Government. Since MITRE resides in a not-for-profit advisory position to their federal sponsors, the research team was able to take an unbiased perspective driven solely by identified issues, the search for improved workflows, and practical opportunities for technology development. The goal of the effort was to inform the US Army community so that it can make responsible, needs-driven decisions regarding gestural interface technologies, and avoid the potential pitfalls that may arise from technology-centered or profit-driven decisions.

The problems focused upon by this research primarily revolved around the collaborative human workflows that occur within Command-and-Control environments. Specifically, the effort targeted US Army-based C2 environments, such as a notional fixed command center, a mobile command center, and the environment of the dismounted soldier in the battlefield. The primary issue is that the currently-implemented technologies, while independently sufficient, present constraints when distributed personnel are collaborating across them. The research team addressed this cross-platform issue by adhering to a Systems Engineering framework that required a holistic approach to the "system" of distributed C2 personnel and their technologies. The goal for the final output was to demonstrate how these technologies may come together as a system to support a more efficient, dynamic, and effective operational workflow than today's reality.

After carefully examining the field of current and emerging gestural interface technologies, and mapping them against available HCI-related research findings, the team concluded that US Army personnel may indeed benefit from effectively and appropriately implemented technologies from this domain. At a high level, gestural technologies offer C2 personnel an ability to conduct more efficient and collaborative workflows across distributed environments. The exact details of these workflows, including the key users, actions, and technology paradigms, are outlined in the content of the whitepaper. In an effort to be as prescriptive as possible, the research team decided that it would be valuable to include a sizable section within the whitepaper dedicated to instructing the user on how to implement gestural

* Approved for Public Release: 10-3988. Distribution Unlimited.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 187–194, 2011.

technologies for C2 application. In this section, they outline the key design patterns to selecting proper solutions and developing effective interaction design frameworks. The nature of this instructional portion ranges from high-level design principles and best practices down to detailed visual demonstrations of recommended gestures.

Keywords: Multi-touch & Tangible User Interfaces, Immersive Computing, Multi-modal Collaboration, Interactive Surface Interface, WIMP Interface, Graphical User Interfaces (GUI), Multi-Touch Surface, Tangible Interfaces, 3D User Interfaces, Wearable Computing.

1 Introduction

In this document we will first discuss a new method for designing the interaction mechanisms of new technologies for the command-and-control (C2) environment. We will then present the resultant operational framework of new concepts considering the environments, workflows, and known human-based issues experienced in C2 environments. We attempt to adequately address a subset of critical C2 issues and clearly present technically-plausible, gesture-based solutions in a novel way that would be understandable and beneficial to military sponsors.

To inspire our design and provide meaningful documentation for our sponsors we chose to create a visual narrative that took the traditional story-board process a step further with flash animation. It is generally accepted knowledge that the visual communication of novel systems and technologies is far more effective than attempting to describe the concepts in language-based detail. Furthermore, when developing new interface concepts, it is not only important to understand the functionality of the product, but to visualize the interactions it promotes. Interaction design is inherently visual, so the means of conveying that design should also be visual. Especially in an instance where the user group and scenario are highly unusual or inaccessible, as the active C2 environment is, visual animation provides an additional level of fidelity not necessarily found in static storyboarding. The visual narrative method allowed us to illustrate a cohesive case study that enabled our audience to understand not only the context in which these technologies could exist, but the possible dynamics, interactions and transitions between them. A primary goal for this project was to demonstrate interactive technologies as a unified system of components that support a collaborating set of users across a range of environments and workflows.

It is essential that we discuss the motivation of this work, as the application of gesture-based technologies to command-and-control is a decision that is certainly up for evaluation. Gestural interfaces, especially multi-touch technologies, have existed for many years. Humans, whether civilians or military personnel, have shown to prefer gestural interfaces because they present an interactive language that naturally maps to their own experiences with the physical world. Gestural interfaces frequently allow for the *direct manipulation* of on-screen information instead of requiring intermediary tools and training. The result is frequently a swifter, more natural, and less error-prone interaction. Due the critical and time-sensitive nature of military operations, where errors and delays are risky and unacceptable, it is essential that we as engineers and designers examine the opportunities for appropriate implementation of gestural interfaces within these environments.

The remainder of this document discusses an example of what we consider to be the appropriate implementation of gestural interfaces to the C2 communication- and work-flow. We have organized our recommendations by operational environments beginning with a discussion of the technologies that appear in a representative Tactical Fixed Center (a fixed command and control center), followed by a notional Tactical Mobile Center (mobile command center frequently deployed into the theatre) and finally, the environment of a notional Dismounted User (soldier on the ground).

2 C2 Environment: Tactical Fixed Command Center

2.1 Overview of Fixed Tactical Center Technologies

Key players in the tactical fixed center include a team of Strategic Planners gathered around a meeting table, and data analysts, including an "Immersive Planner", at remote and/or independent work stations.

The Tactical Fixed Center receives guidance via video conference with the Strategic Command Center (a third entity – not considered in our recommendations – charged with setting the objective of the mission) that there is an opportunity to take control of a critical enemy location. The guidance is packaged in a dynamic digital plan that can be loaded onto a tabletop touchscreen surface when authorized strategic planning personnel initiate a session. Key personnel in the Tactical Fixed Center gather around the planning surface to view the mission objectives and imagery sent in the digital plan. Users then select and manipulate additional relevant mission and environmental data supplied and supported by the data analysts. Strategic Planners interact with the table via haptic inputs and by manipulating tangible objects representing different types of data and controls placed directly on the surface. This is heavily collaborative as experienced personnel with varying fields of expertise and motivation need to analyze, discuss, and agree on a proper course of action.

As the strategic personnel are discussing a potential plan, they realize that they need real-time ground-level clarification of a particular geographic area to understand the potential hazards and feasibility of the plan. They submit a request for information by identifying the area in question on the map and assigning it to the "immersive planner".

Fig. 1. Strategic Planning Personnel and Interactive Mission Planning Table

The Immersive Planner is a newly conceived role providing a data analyst the ability to interact with ground data in an immersive digital environment. The displays feature a virtual environment created from available mission-related and environmental data. The strategic planners' request from the planning table appears on the center display and a video chat window opens to allow for direct communication with the Strategic Planners. Using hand gestures, the Immersive Planner is able to navigate through the virtual environment specified by the strategic team and communicate her observations by flagging important areas. These areas of interest are then mapped back to the display on the Strategic Planning Table. The Immersive Planner can further detail her findings by chatting with the Strategic Planners in real time.

The Immersive Planner's location, findings, and corresponding environmental information (weather data, dismounted soldiers, etc.) can all be layered and displayed on the planning surface's map to provide situational awareness to the Strategists as they design the plan. During this activity, an ad hoc request may come in from the theater requesting real time support in an unknown area. The Strategic Planners may override the location of the Immersive Planner and "move" them to area of interest in their virtual environment to explore the area for potential hazards based on (close to) real-time video feeds. They may then relay their findings back to the Strategic Planners, or where applicable, directly to the dismounted soldiers.

With the insight from the Immersive Planner, the Strategic Planners are able to come to a consensus around an approved course of action. Documentation of the plan collected and recorded on the surface is sent forward to the Tactical Mobile. This allows the Mobile Center, as an outpost of the Fixed Center, to have access to the same information and situational awareness as the Fixed Center and Dismounted Soldier. We will explore this role further in subsequent sections.

2.2 Discussion

We believe that the Fixed Tactical Command Center can benefit from the implementation of innovative technologies that specifically promote *agile collaboration*. As demonstrated in the vignette above, we have chosen to showcase a combination of different types of technologies to accomplish this objective. In this section we will discuss each medium – the multi-touch surface, the Tangible User Interfaces (TUI), and the Gestural Interfaces – independently to develop an understanding of each technology and their capabilities.

It is not by chance that we decided upon a collaborative planning surface to demonstrate the technologies in this section. We want to find an analogy for the common operational task of personnel gathered around a table formulating a course of action because we wanted a medium that would align with known traditional workflows. We chose to exemplify multi-touch, TUI, and gestural technologies where they enhanced common workflows, behaviors, and expectations.

3 C2 Environment: Tactical Mobile Center

3.1 Overview

Four soldiers are stationed in a notional Mobile Tactical Operations Center. They receive a message from the Fixed Tactical Command Center regarding a mission update. The objective of the team in the mobile center is to augment and support the flow of communication and guidance between the strategic planners in the Fixed Tactical Command Center and the Dismounted Soldiers in the field. They also control the coverage of a small unmanned aerial system (UAS) which can provide real-information to the mission.

Before the message arrived, three of the four soldiers in the vehicle were monitoring the operational picture on tablet displays within the vehicle. The displays are rugged tablet devices that can be removed and re-mounted in any of the various docking stations throughout the vehicle. Docking locations include the passenger dashboard, the sun visors, the back sides of the front seats, and even on the outside of the vehicle. The visual interface of the tablets is similar to the software being run on the table in the Fixed Center (aside from the lack of tangible objects). This means that users in the Mobile Center have much of the same information available to them in the same formats as in the Fixed Center, which facilitates collaboration and minimizes misunderstandings.

One soldier in the Mobile Center is piloting a small UAS that was launched from the vehicle earlier in the day. He sits in the specially-equipped passenger seat featuring a dashboard-mounted console for controlling and monitoring small UAS's. The console also provides a location for the soldier's tablet to maintain situational awareness as he is piloting the UAS.

Once a plan has been distributed to dismounted soldiers (see next section), the Mobile Center acts as a conduit between the Strategic planners and those executing the plan. As information brokers between the two entities, they are responsible for ensuring that the dismounted soldiers have the right information uploaded to the forearm-mounted display, and that the Fixed Center is kept abreast of the events in the theater so they will be prepared to provide assistance should it be needed.

3.2 Overview of Tactical Mobile Center Technologies

While the technologies employed in the notional Mobile Center are very different from those seen in the Fixed Center, they capitalize on many of the same interaction techniques. In the following section we will describe how recommendations for tools used in the Mobile Tactical Center can use the same interaction techniques. Because we wanted the design guidelines presented in the Fixed Center to be a reference for those mediums regardless of application, we will only contribute design suggestions where there are completely novel interactions to be discussed.

The soldiers in the notional Tactical Mobile Center use personalized and rugged haptic tablet computers featuring multi-touch interfaces. As explained in the overview of multi-touch technologies in a previous section, this medium was chosen to promote efficiency and enable agile collaboration (for a detailed explanation of this, see the previous section outlining multi-touch technologies). Multi-touch interfaces are a particularly intuitive medium for viewing, interacting with, and manipulating data. On the basis of its inherent usability merits alone, it is recommended that the soldiers in the Mobile Tactical Center have access to multi-touch technologies in an adaptable form factor. Tactically manipulating data directly instead of using the metaphorical proxy of the mouse minimizes mistakes and time wasted in the physically constrained, high-risk, high-stress environment of the Tactical Mobile Center.

4 C2 Environment: Field of Battle

4.1 Overview

The final environment in our workflow features a notional dismounted soldier in the field of battle utilizing gestural technologies. The priorities of this position are clear: optimize situational awareness for the Fixed and Mobile Centers, avoid threats, and complete the operational goal within the necessary bounds. The dismounted soldier workflow is high-risk and time-critical, so the technology designed must be as unobtrusive and efficient as possible.

Fig. 2. Images from Field of Battle environment

During our concept animation, we see the Dismounted Soldier initiate a request for Close Air Support. To do this, he activates the arm-mounted display by removing the protective cover. He first sees a main menu that features the ability to create a request, leave a visual tag, see the mission plan, or place a voice call directly to the Tactical Fixed Command Center. In this case, he makes a request to initiate a Close Air Support sequence. This action launches a map view that contains the updated target set and the required path of entry which was determined by the Tactical Fixed Center, including the Immersive Planner.

If he chooses to, the Dismounted Soldier may want to utilize the heads-up display attached to his helmet to receive an "augmented reality" overlay of the mission plan or even just to receive visual indication of alerts or changes in plans. The Dismounted Soldier would need the ability to just "flick on" the mission plan for a brief second in order to aid in decision-making, and then quickly turn it back off to avoid potential distraction. The content on the HUD could even be a semi-transparent, in-context overlay of the Immersive Planner's virtual walkthrough. This could help to guide the

Dismounted User in times of urgency or distress because it would not require the cognitive task of translating directions into actions.

Various lower-ranking dismounted soldiers may wear a small wrist-mounted device that looks similar to an interactive watch. However, this device is also a simple communication mechanism that can display critical and succinct pieces of information, such as a warning icon, a "go ahead" icon, etc. It also allows the soldier wearing it to send out calls for help (e.g. "danger!") or simply tag his or her location. This device would also feature tactile feedback to grab the attention of the wearer when necessary. In an effort to support open communication back to the Tactical Fixed Center, this wrist-mounted mechanism would allow soldiers to directly add geolocated military iconography to command center software with the simple click of a button.

Finally, the Dismounted Soldier may carry an "augmented reality flashlight", which would be a data-enabled flashlight device that would allow a user to project mission data onto any surface in the environment for single-point overlay of contextual information. At appropriate times, this may be less invasive than a HUD as the data would be projected onto the target, instead of being displayed inches from the user's face.

4.2 Overview of Dismounted Soldier Technologies

The primary means of communication for the Dismounted User group leader in our scenario is a flexible interactive display integrated into a fitted forearm sleeve of their uniform. The lightweight sleeve features a protective flap that can be fastened over the display to prevent damage caused by battlefield elements, such as dust and dirt. The device gives the soldier the ability to view the latest course of action, receive and send messages, request guidance or support, leave digital "tags" in the environment, and connect via voice to necessary personnel and environments. The device also exchanges data with external devices, such as laser range finders and GPS mechanisms. Closing the flap prompts the device into a passive "sleep" mode, since battery power is likely at a premium. Of course, this device is simply conceptual, but is represents a possible mechanism supporting the Dismounted Soldier's workflow in a convenient and intuitive manner.

Other dismounted soldiers in the environment wear a small "single point" wrist-mounted device whose form factor is similar to a watch. This device is a simple communication device that can display critical and succinct pieces of information such as a warning icon, a "go ahead" icon, etc. It also allows the soldier wearing it to send out calls for help (e.g. "danger!") or tag his or her location. This device would incorporate tactile feedback to grab the attention of the wearer when necessary.

The Dismounted Soldier may want to utilize the heads-up display attached to his helmet to receive an "augmented reality" overlay of the mission plan or even just to receive visual indication of alerts or changes in plans. The Dismounted Soldier would need the ability to just "flick on" the mission plan for a brief second in order to aid in decision-making, and then quickly turn it back off to avoid potential distraction. The content on the HUD could even be a semi-transparent, in-context overlay of the Immersive Planner's virtual walkthrough. This would help to guide the Dismounted

User in times of urgency or distress because it would not require the cognitive task of translating directions into actions.

5 Conclusion

In the past decade, the emergence of Apple's iPod/iPhone product lines and the Nintendo Wii have propelled gestural interfaces to be commonplace tools used in many homes in this country. While the usage of these consumer products may appear unrelated to the needs of military personnel, one must consider the reasons people choose products employing these technologies over others. People prefer gestural interfaces because they present an interactive language that naturally maps to their own experiences with the physical world. If a user wants to select an object with a traditional computing device (e.g. mouse and keyboard), they have to manage a set of cognitive and physical "rules" for the relationships between their desired action and their hands, mouse, keyboard, and objects on the screen. With a multi-touch or gesture-based interface, the user interacts directly with their desired object or action – a perfect correlation to real world interaction. The result is a swifter, more natural, and less error-prone interaction. It is this fact that brings us back to our exploration of gestural technologies for command-and-control environments. Due the critical and time-sensitive nature of military operations, where errors and delays are risky and unacceptable, it is essential that we as engineers and designers examine the opportunities for appropriate implementation of gestural interfaces within these environments. We hope this document addresses those opportunities and provides valuable guidance and recommendations for the implementation of these technologies.

Talking to Strangers: Using Large Public Displays to Facilitate Social Interaction

Elisa Rubegni[1], Nemanja Memarovic[2], and Marc Langheinrich[2]

[1] Faculty of Communication Sciences
[2] Faculty of Informatics
University of Lugano, 6904 Lugano, Switzerland
{elisa.rubegni,nemanja.memarovic,marc.langheinrich}@usi.ch

Abstract. Alumni events and homecomings provide opportunities to reconnect and reminiscence with old friends and colleagues, i.e., they aim to reinforce connections between community members. Although these events explicitly foster social interaction, the first step in engaging with others can still be difficult. To help "break the ice", we have built *USIAlumni Faces*, a 'yearbook' application running on a public display that is operated via a gesture interface. We deployed USIAlumni Faces at a large university alumni event, which gave us the opportunity to observe and understand learning techniques for gesture interfaces and their role in supporting the emergence of social interaction in public spaces. We found that gesture-based interfaces support the natural diffusion of interaction patterns in public spaces through the observe-and-learn model, and that sensory-motor patterns can aid social interaction in public, as they act as conversation starters between both strangers and acquaintances.

Keywords: public displays, gesture interfaces, social learning, interaction design.

1 Introduction

Alumni events and homecomings are important part of the university life: they provide opportunities to see and talk to old schoolmates and reinforce the connections with people that we have not seen in a long time. Most of the people catch up on the news and reminiscence 'the good old days', i.e., joint experiences created while they studied. These experiences are tied to different communities or social groups: some of the people took classes together; some of them were part of the university's sports team (e.g., soccer, basketball, or volleyball); some of them were part of a student organization; and some of them simply used to 'hang out' and go out on the weekends. The network of people we meet while studying is tremendous.

At the alumni events all of these people come together. Although these events are highly communal in their nature and stimulate communication and socializing, the first step in engaging in a conversation can be difficult: some people are shy and intimidated by social embarrassment, sometimes it is hard to recognize old friends, while at other times people's interests have changed and it is hard to find a common topic. In these circumstances, yearbooks can be a great aid for remembering the past

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 195–204, 2011.

school days: people can recognize their friends from the time when they studied together, they can see the list of student organizations as well as sports teams, and they can find information about classes, projects, and other educational aspects.

The yearbook metaphor was an inspiration for the *USIAlumni Faces* installation, an interactive yearbook application running on a large public display and operated via a gesture interface. For the simplicity of the installation the yearbook application contained only the most important part of the yearbook, i.e., images of people, and did not include various lists (sports teams, student organization, classes, projects etc.). The installation was built to serve as the conversational 'ice breaker' by stimulating discussion around the presented content and by offering an interaction modality that makes user's actions publicly visible through the gesture-based interface and a large public screen.

The USIAlumni Faces was deployed at a large university alumni event in Lugano, Switzerland. At the event we observed and video recorded people interacting with the installation. On-the-spot observations and in-depth video analysis allowed us to identify a learning technique for gesture interfaces as well as their role in supporting the emergence of social interaction in public spaces.

This paper is structured as follows: first we introduce related work on social interactions around public displays, direct manipulation interfaces, and large-screen collaboration. After that we describe the USIAlumni Faces installation and the deployment setting. We present our findings from the observations and video analysis followed by a discussion on the natural spread of gesture interfaces through the *observe-and-learn model,* as well as their role as a conversation catalyst via *sensory-motor patterns*. We close with conclusion from our findings.

2 Related Work

Our work intersects several active research areas, most notable social interaction around public displays, direct manipulation interfaces, and large-screen collaboration. We will briefly summarize related work in these fields in turn.

Public display systems have been shown to be an effective means to deploy situated social software 1, i.e., software systems that are designed for a specific community or social group. An early example of such situated social software is the Groupcast system by McCarthy et al. 2. Groupcast allowed users to upload profiles that reflected their interests within a working environment. After identifying users that were standing in front of a large public display via infrared badges, Groupcast would then show common interests of the people standing in front of it.[1] Public displays have also been used in semi-public events such as academic and industry conferences – events that have a similar setting to alumni events. McDonald et al. 3 developed three applications to help socializing at scientific conferences: Auto-SpeakerID, an application that would show the name and affiliation of a person asking a question on a large public display; Ticket2Talk, which showed the name, affiliation, and a user-chosen image representing his or her interests, whenever a

[1] While the system initially showed the intersection of interests, it quickly became apparent that the *union* of interests was significantly more effective at starting conversations.

delegate would pass a large public display; and NeighborhoodWindow, an application that showed both the intersection and union of peoples' interests on a large public screen, similar to Groupcast. The authors report that these applications did increase the sense of a community among the attendees.

Translating traditional, non-digital tools for social interaction into digital counterparts has also been shown to be an effecting approach for supporting communities. Churchill et al. 4 investigated the properties of paper-based notice boards before they built PlasmaPoster, a digital workplace notice board for stimulating serendipitous social interaction. PlasmaPoster posts were based both on user-contributed material, as well as on automatically downloaded/streamed web content. A large majority of the users found the content interesting and began conversations around it. We recently adopted a similar approach for designing a digital public notice area system 5.

While many prior systems focused on private and semi-public environments (e.g., alumni events) [2, 3, 4], Peltonen et al. investigated how people reacted to large displays in a public urban environment 6. They deployed CityWall, a large public display with multitouch support in the center of Helsinki, Finland. CityWall simply displayed random Flickr images that were tagged with 'Helsinki', and let users browse through them. An 8-day trial revealed that the support for parallel interaction would repeatedly prompt strangers to interact with each other, as their image manipulations often "spilled over" into another users part of the screen.

USIAlumni Faces builds on the above work and investigates social interaction around a large public display that involves neither colleagues nor strangers, but past friends and acquaintances. USIAlumni Faces also uses a novel interaction modality, a toy flashlight that acts as the main input device, thus opening the screen for onlookers while focusing control on a single artifact.

The USIAlumni Faces interaction model is based on the direct physical interaction of digital media (students' pictures) embedded in the physical environment (a large public screen). There is a growing recognition of the benefit of physical interaction, as it enables new form of experiential learning and affords collaborative interactions 8. Marshall 9 argues that the rich physical experience provided by the direct physical manipulation of objects is key to intellectual development. A growing number of studies (e.g., Kolb et al.) investigated the capability of direct manipulation interfaces in supporting both individual and collaborative activities. User engagement has been found to be raised by the tactile experience provided by a touch screen (e.g. Jacucci et al. 10, Kierkels and van den Hoven 11) and by tangible objects used in an interactive surface (e.g. Jordà et al. 12). However, while evaluation studies of specific interfaces have recognized the advantages of various interactive tools on supporting human activities, a theoretical understanding of the psychology of interactivity 13 is still missing. Also, it is quite difficult to understand whether the interactivity is a value per se, or whether it can actually support the comprehension of contents.

A few studies go in the direction of demonstrating the benefits or the disadvantages of a specific Tangible User Interface (TUI) technology compared to other interaction modalities (mouse-based, multi-touch, analogical physical interaction) or interface styles (e.g. GUI). For instances, TUIs, physical (traditional) modality and a Graphic User Interfaces (GUI) are compared for understanding the level of engagement of children in doing a jigsaw puzzle 14. In other cases, the assessment is focused on the user performance evaluation through the comparison of TUI and multi-touch interfaces (e.g. Lucchi et al. 15), or TUI, multi-touch, and mouse (e.g. Tuddenham et al. 16).

Our research within the USIAlumni Faces case study is focused on understanding the role of direct manipulation for supporting a "fluid" 17 and engaging interaction with contents, with the purpose of affording social and collaborative behaviors. The direct manipulation of digital media is claimed to be engaging since it enables "natural" interaction through the use of everyday objects 18. When objects and the actions connected to them are meaningful for users, technology becomes transparent and the interaction natural. Ideally, interaction is based on patterns that are evocative or denotative of the contents or effects, allowing people to not reflect on the medium they are using but instead focus on the content 19. Furthermore, gesture-based patterns make actions visible to both users and bystanders, thus improving mutual awareness and consequently the possibility for people to understand the activity of others 20.

Last but not least, USIAlumni Faces relates to previous work on large-screen collaboration, demonstrating turn-taking and shared learning patterns. Russel et al. 21 found that their touch-based (single touch), large public display application BlueBoard for workplace discussions and content sharing would naturally provide learnability through observable interactions, as well as emerging and fluid control through direct social interaction. Rogers and Lindley 22 found that screen orientation of a large display significantly affected collaboration in a workgroup setting: horizontal orientation encouraged awareness and collaboration while vertical setups stifled exchange. While USIAlumni Faces uses a vertical orientation, its setting is significantly different from Rogers and Lindley, who only looked at small groups that were tasked with collaboratively solving particular goals. Having a vertical orientation of the screen allowed us to draw in larger groups of people to share the display. Ha et al. 23 looked at the implications of different input devices on the interaction around a common tabletop application. While indirect input devices such as mice lowered the physical effort required and were more familiar to users, direct manipulation instruments such as styli offered noticeable gestures that made intentions more apparent, offered better coordination in joint tasks, and supported more fluid gestures. USIAlumni Faces direct interaction control through a toy torch explicitly supports learnability and openness through its visible gestures, thus aiding our goals of stimulating social interaction.

3 The USIAlumni Faces Yearbook Application

As part of a university alumni event, we built and deployed an interactive installation called *USIAlumni Faces*, which projected a virtual "yearbook" (i.e., photos of the alumni organized by year and faculty) onto a large public screen (cf. Figure 1). To navigate and browse through the yearbook, participants had to perform a 'page flip' gesture with a custom-built input device – a Wii remote control and an infrared pen hidden inside a toy torch casing. The installation is perceived as an *interactive artifact* that acts as a cultural mediator 24: the process of learning involves a subject (the learner), an object (the task or activity) and a cultural artifact. The interactive installation mediates the relationship between the subject and the object of any activity. Thus, the interaction with the artifact encourages the negotiation of meaning among the learners and, consequently, stimulates the learning process itself.

The artifact design followed a co-evolutionary process in which concept, technology, and activity design were carried out simultaneously so that each strand of the process informed the other. The gesture-based interaction model was defined in a laboratory setting, in which the most meaningful mapping of input actions (gesture-based) and output responses (visual-based) was assessed through several user trials.

The event provided a unique opportunity to observe and understand learning techniques for gesture interfaces, as well as their role in supporting the appearance of social interaction in public spaces. Over 200 people used the artifact during the event, which took place on a single day. One of the researchers introduced the system and explained its purpose, but no explanations were given on how to interact with the artifact. Participants then freely explored the interface in order to understand the interaction model. The researcher notified participants that their interactions with the artifact were observed and videotaped for later analysis.

Fig. 1. A group of visitors interacting with the USIAlumni Faces

4 Findings

The interaction modality was spread through imitation: people learnt from each other how to interact with the artifact. The process of imitative learning is well known in psychology 25: the observer attempts to copy the behavioural strategy of the other and to reproduce the intentional actions of the other, including the goal toward which they are aimed 26.

People looking at the others playing with the artifact internalized the interaction modality and then customized it: in a few cases we observed people created their own strategies for interacting with the artifact (such as click on the right/ left angle or making a short and quick gesture for flipping the page). There were also cases where some of the users, incidentally, discovered new interaction modalities (e.g. click in the middle of the page).

Just in few cases a new user asked for an explanation on how to interact with the artifact. In those cases either other people showed the interaction modality by performing the gesture, or in cases when no other people were around the research intervened to show the pattern without giving any oral explanation.

During the event small and big group of people stepped in front of the artifact: the individual interaction was rare while the interaction of group (from 2 to 8 people) was that more usual. In a few cases two people used the input device together (Figure 2), but majority of interaction with the device is individual while the others suggest the information to be looked for.

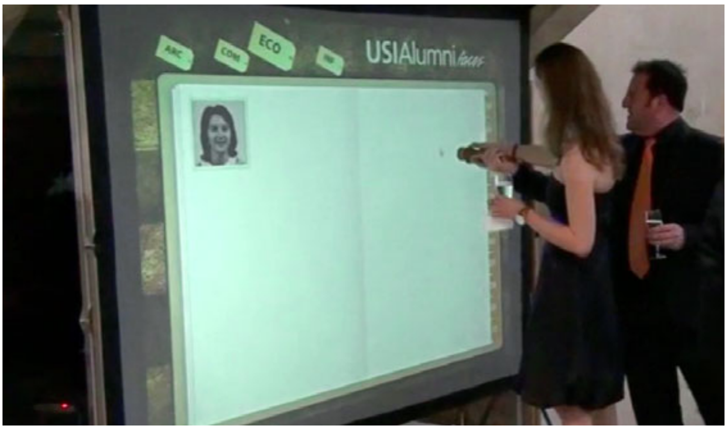


Fig. 2. Two people use the input device together

Also people ‘passed’ the interaction from one another in turns. In average, one interaction lasted from two to ten minutes.

An interesting element that demonstrated the attendants’ wish for sharing was “pointing” (Figure 3): people indicated, using their fingers, the picture/s on the screen and commented with others (in many cases they laughed!). Indeed, someone touched the screen hoping that they would magnify the picture.

The artifact demonstrated not only that interaction patterns were easy to understand and mimic, but also that they could stimulate social interaction. In particular, we observed that during the whole event, people who met in front of the artifact have started a conversation; in many cases, they continued also after the stopped interacting with it. We observed multiple occasions where people that never met during the university started talking in front of the artifact and had the opportunity to introduce each other. Unexpectedly, pictures also acted as memory aids: they allowed people who had not seen each other since graduation to recognize the person standing next to them through images displayed on the screen. The images also stimulated people bringing up memories and stories from the past: they began to tell interesting anecdotes about their life at USI. Often people engaged in the conversation called their friends to join them. In one interesting case a girl took a picture of the displayed images.

Fig. 3. A group of people pointing and talking while playing with USIAlumni Faces

Although the artifact was mainly designed for the adults it was also engaging for children. A couple of children approached the device and started playing with it; they were very absorbed by the interaction modality not really by the contents even if the pictures fascinated them. One of them in a few occasions asked the mother to pick her up in neck to change the faculty and the years.

5 Discussion

Overall, the concept highly motivated people to use the artifact: people enjoyed looking at their own pictures, and those of their classmates, from their first year in University. The artifact was also a catalyst that encouraged and animated both strangers and acquaintances to start a conversation, and in many cases, to reminisce together about the good times they had at the University. Our findings are organized around two key results: 1) Public spaces support the *natural diffusion* of gesture-based interaction interfaces through the observe-and-learn model; and 2) Sensory-motor patterns aid *social interaction in public*, as they act as conversation starters between both strangers and acquaintances.

Natural Interaction Diffusion: Gestures can enrich the user experience by creating an additional level of interest and intuitiveness in the way a user can control and interact with a system 20. Additionally, gesture interface in public spaces also support the ability to “diffuse” the interaction technique to bystanders through an observe-and-learn model. Our analysis of people’s behavior during the event showed that the gesture-based model increased the visibility of actions, and that it supported the understanding of the user’s intention in performing the action. People who observed the interaction of others were subsequently able to learn the interaction technique with ease, a process of *natural diffusion*.

Social interaction through sensory-motor patterns: The observe-and-learn model described above not only made it easier for people to learn the gesture-based interface, but it also encouraged spontaneous interaction among attendees. The sensory-motor patterns offered by the artifact gave users the opportunity to share their intentions. Some users, e.g., used excessively large gestures with the input device, even though small movements sufficed, in order to signal their openness for social contact.

The artifact also stimulated *social debate*, as well as *collective usage*. People tried to interact collaboratively, e.g., one user flipped the page using the torch while another tried to zoom using his hands (which wasn't supported by the interface, however). In many cases, two people used the torch together by either repeatedly passing it between each other, or by grabbing and moving the torch holder's hands. This confirms previous findings, e.g., from Brignull et al. 27 and Peltonen et al. 6, who observed similar *teacher-apprentice* relations between collaborative users of a shared interactive displays.

Our choice of a yearbook application also confirms some of the findings around the social use of photo-sharing applications, e.g., by Taylor and Cheverst 28, which proved to be a good tool for strengthening or re-connecting social relationships. During the alumni event, we observed several such "re-connects" between old friends in front of the display. We are currently evaluating a number of additional deployment options of similar public display installations, e.g., in public parks, in order to further explore the role of public displays in fostering social interactions.

6 Conclusion

Homecomings and alumni events enable schoolmates and colleagues to reestablish their connections and reminiscence 'the good old times'. Although these events are highly social the initial step in starting a conversation can be difficult since most of the people change after graduating (e.g., physically appearance, change of interest). To ease the starting step we have built USIAlumni Faces, an interactive installation that serves as an 'ice breaker' by displaying a 'yearbook' that is operated through a gesture interface. The interactive artifact was deployed at a large alumni event where we observed and video recorded people interacting with it. Our initial findings from the observations and video analysis reveal the natural diffusion of gesture patterns in public spaces, i.e., people were able to learn the interaction modality through the observe-and-learn model. The observe-and-learn model also acted as a social catalyst that sparked conversation and discussion among people. To further verify our findings we are planning to adapt the artifact and deploy it in several public spaces, e.g., public parks, bars, and community centers.

References

1. Shirky, C.: Situated Software, Networks, Economics and Culture mailing list. First published on (March 30, 2004), http://www.shirky.com/writings/situated_software.html
2. McCarthy, J., Costa, T., Liongosari, E.: Unicast, outcast & groupcast: Three steps toward ubiquitous, peripheral displays. In: Gregory, A.D., Brumitt, B., Steven, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 332–345. Springer, Heidelberg (2001)
3. McDonald, D.W., McCarthy, J.F., Soroczak, S., Nguyen, D.H., Rashid, D.H.: Proactive displays: Supporting awareness in fluid social environments. ACM Transactions Computer-Human Interaction 14, 4, 16, 1–31 (2008)

4. Churchill, E.F., Nelson, L., Denoue, L.: Multimedia fliers: Information sharing with digital community bulletin boards. In: Proceedings of the First International Conference on Communities and Technologies, pp. 97–118. Springer, Heidelberg (2003)
5. Alt, F., Memarovic, N., Elhart, I., Bial, D., Schmidt, A., Langheinrich, M., Harboe, G., Huang, E., Scipioni, M.P.: Designing Shared Public Display Networks — Implications from Today's Paper-Based Notice Areas. In: Proceedings of the 9th International Conference on Pervasive Computing (to appear, 2011)
6. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., Saarikko, P.: "It's Mine, Don't Touch!" Interactions at a large multi-touch display in a city centre. In: Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 1285–1294. ACM, New York (2008)
7. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development. Prentice-Hall Inc., New Jersey (1984)
8. Hornecker, E.: Graspable Interfaces as Tool for Cooperative Modelling. In: Proceedings of the 24th Information Systems Research Seminar in Scandinavia, vol. 3, pp. 215–228 (2001)
9. Marshall, P., Cheng, P.C., Luckin, R.: Tangibles in the balance: a discovery learning task with physical or graphical materials. In: Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction, pp. 153–160. ACM, New York (2010)
10. Jacucci, G., Morrison, A., Richard, G.T., Kleimola, J., Peltonen, P., Parisi, L., Laitinen, T.: Worlds of information: designing for engagement at a public multi-touch display. In: Proceedings of the 28th annual SIGCHI Conference on Human Factors in Computing Systems, pp. 2267–2276. ACM, New York (2010)
11. Kierkels, J., van den Hoven, E.: Children's haptic experiences of tangible artifacts varying in hardness. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges, vol. 358, pp. 221–228. ACM, New York (2008)
12. Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction, pp. 139–146. ACM, New York (2007)
13. Sundar, S.S., Xu, Q., Bellur, S.: Designing interactivity in media interfaces: a communications perspective. In: Proceedings of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 2247–2256. ACM, New York (2010)
14. Xie, L., Antle, A.N., Motamedi, N.: Are tangibles more fun?: comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, pp. 191–198. ACM, New York (2008)
15. Lucchi, A., Jermann, P., Zufferey, G., Dillenbourg, P.: An empirical evaluation of touch and tangible interfaces for tabletop displays. In: Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction, pp. 177–184. ACM, New York (2010)
16. Tuddenham, P., Kirk, D., Izadi, S.: Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In: Proceedings of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 2223–2232. ACM, New York (2010)
17. Maes, P.: Fluid Interfaces Group, Massachusetts Institute of Technology Media Lab, http://fluid.media.mit.edu/people/pattie/about/pattie.html
18. Dourish, P.: Where the action is: the foundations of embodied interaction. MIT Press, Cambridge (2001)

19. Rizzo, A., Rubegni, E., Caporali, M.: Why here and now. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5615, pp. 729–737. Springer, Heidelberg (2009)
20. Milekic, S.: Towards Tangible Virtualities: Tangialities published. In: Museums & Web 2002: Selected Papers from an International Conference, pp. 189–196. Archives & Museum Informatics (2002)
21. Russel, D.M., Drews, C.: Social Aspects of Using Large Public Interactive Displays for Collaboration. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, pp. 663–670. Springer, Heidelberg (2002)
22. Rogers, Y., Lindley, S.: Collaborating around vertical and horizontal large interactive displays: which way is best? Interacting with Computers 16(6), 1133–1152 (2004)
23. Ha, V., Inkpen, K.M., Whalen, T., Mandryk, R.L.: Direct Intentions: The Effects of Input Devices on Collaboration around a Tabletop Display. In: First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp. 177–184. IEE Press, New York (2006)
24. Vygotsky, L.S.: Interaction between learning and development. In: Mind in society: The development of higher psychological processes, pp. 79–91. Harvard University Press, Cambridge (1978)
25. Tomasello, M.: The cultural ecology of young children's interactions with objects and artifacts. In: Winograd, E., Fivush, R. (eds.) Ecological Approaches to Cognition: Essays in Honor of Ulric Neisser, pp. 153–170. Lawrence Erlbaum Associates Inc., Mahwah (1999)
26. Rizzo, A., Del Monte, M., Rubegni, E., Torsi, S.: The interplay between sensory-motor and intentional affordances. In: Children and embodied interaction workshop: seeking common ground. Position Paper at the Workshop Embodied Interaction at IDC 2009, Como (2009)
27. Brignull, H., Izadi, G., Fitzpatrick, G., Rogers, Y., Rodden, T.: The introduction of a shared interactive surface into a communal space. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 49–58. ACM, New York (2004)
28. Taylor, N., Cheverst, K.: Social interaction around a rural community photo display. Int. J. of Hum.-Comp. Studies 67(12), 1037–1047 (2009)

The Grid Intelligent Planning Framework: Planning Electric Utility Investments in a Time of Accelerating Change

Geoff Ryder, Fatimah Shahid, and Sui Yan

SAP Labs LLC, 3412 Hillview Drive,
Palo Alto, CA 94304, USA
{geoff.ryder,fatimah.shahid,sui.yan}@sap.com

Abstract. Over the next ten years, electric power utilities will be required to invest billions of dollars to meet public policy goals for a greener, smarter electricity grid. Renewable generation portfolio standards, electric vehicle infrastructure, advanced metering infrastructure, and the replacement of aging grid assets are some of the factors driving these new investments. The Grid Intelligent Planning Framework using GridLAB-D is an advanced forecasting solution that allows utility business and engineering experts to collaborate on forecasting models, and thereby to reduce the time needed for a capital investment planning cycle. This solution facilitates wise and timely investment decisions as the pace of change accelerates in the electric power industry.

Keywords: energy management, forecasting, risk analysis, capital investment, collaboration, electric power utilities, visualization.

1 Introduction—Planning Roles and Responsibilities

Electric power utility companies face profound challenges as they modernize their operations for the demands of the 21st Century. Some challenges affect the industry on a world-wide scale, such as the imperative to respond to climate change [1]; and some are specific to particular regions, such as the need to replace aging equipment [2]. In response to these challenges, significant new capital investments in generation, transmission, and distribution-level infrastructure will be required [3]. To make good decisions, utility planners build models of their operations that allow the impacts of future investments to be simulated. In this environment, planners need to build high quality simulations quickly that lead to sound investment forecasts. Accompanying this, we see a need for new solutions that speed up the business process of creating quality simulations [4].

Here we describe a software solution called the Grid Intelligent Planning Framework, or “GridIntell”, that has the potential to achieve faster results. The human-computer interaction components of the GridIntell project are currently under development, and the work presented here reflects early user experience research and prototype development.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 205–214, 2011.

As an introduction to the subject, consider the roles and responsibilities involved in producing a capital investment forecast. Table 1 describes key roles we identified through market research and stakeholder interviews. From demographic research on each role, we discovered what collaboration patterns, tools, and experience levels were commonly required to build forecasts. The roles fall roughly into three functional teams: an economic team (data analysts, energy economists), an engineering team (power engineers), and a management or supervisory team (business analysts, senior planning engineers, and on up to C-level executives). We verified that the end-to-end process is typically led by a senior planning engineer, who articulates the goals for the three teams, guides them in their work, and presents investment options to executives based on the forecasting results.

Table 1. Responsibilities and selected pain points for participants in capital investment planning processes at utility companies

Title	Responsibility	Pain Points
Energy Economists	·Select economic parameters for simulation models; quantify model risks	·Key demographic information changes faster than the forecasting project cycle
Data Analysts	·Collect economic data; enter parameters and run simulations to support forecasts	·Lack of comprehensive and easy-to-use modeling tools
Power Engineers	·Builds models; simulates and recommends grid infrastructure upgrades	·Lack of comprehensive and easy-to-use modeling tools
Business Analysts	·Incorporate forecasting results into planning recommendations for CEO/CFO	·Lack of timely support for generating forecasts
Senior Planning Engineer	·Manages economics and engineering teams who build simulation models, and forecasts	·Difficult for teams to coordinate model updates given new information
CEO/CFO	·Set and review company financial goals	·Lack of evidence to support business decisions

A common theme among the pain points we observed was an absence of comprehensive, collaborative models, forcing the teams to work independently, and to then throw their results "over the wall" to the next team. That process in turn led to the senior planning engineer's pain point—an inability to quickly coordinate comprehensive model updates given new information—and thus led to suboptimal forecasts for the management team to review. Further research revealed unexplored avenues for collaboration that could address these pain points, and which we will describe in the following sections. Our new approach provides members of the different teams more useful touch points across the entire forecasting process.

2 Value Proposition of the Grid Intelligent Planning Framework for Electric Power Utilities

In his 2010 paper analyzing a collection of large scale energy modeling scenarios, Keles discussed the fact that it is extremely difficult to judge their robustness [5]. They depend on such a large number of unknown input parameters—prices of fuel inputs, generation technology, health of grid assets, population growth, business activity, energy efficiency of appliances, consumer behavior—that there is a large degree of subjective choice involved in every scenario design. In this situation, he finds that the best forecasting scenarios are achieved when the designers debate the model assumptions, change the parameters frequently, and rerun the models many times.

Value Proposition: *The Grid Intelligent Planning (GridIntell) Framework facilitates frequent model updates, as well as close collaboration among experts from different domains, resulting in what we believe will be higher quality forecasts from better-designed scenarios in a shorter time.*

Fig. 1 illustrates the process. Here, a cross-functional planning team at a utility company has been tasked with developing forecasts that will drive future capital investment projects. The design of a forecasting scenario starts at the top left, as the economics team determines demographic and business trends that affect the future demand for electricity. After collecting the necessary data, the economists may use tools from statistical inference, such as hierarchical regression models, to estimate future demand. They may also run discrete event simulation models to estimate and verify properties of complex economic systems for which closed-form calculations are not practical.

Second, after the forecasted economic parameters are ready, the engineering team uses them to guide their models of the utility's power grid assets. New economic parameters lead to an altered distribution of electric loads around the service territory, and the engineers develop physical models of the grid to predict where new investments in equipment are needed. Technical software—only accessible to specialists—is used in this step to compute quantities such as the real-time power flow through simulations of the grid (see [3], [6]).

Third, the combined economic and engineering results are turned over to a managerial decision maker to analyze. New investment needs suggested during the planning process must be reconciled with budget and regulatory constraints. While we expect the economics and engineering teams to try to take the constraints into account in advance, the scope of needed investments in future years will likely be so large as to cause constraint violations. The response at this step—to push for larger capital budgets, for regulatory changes, or for better options from the economics and engineering teams—must come from a management perspective.

We describe the serial progression through these three steps as a **long cycle**, and this could take days, weeks, months, or even years depending on the scope of the planning effort. Each functional team must finish before allowing the next team to begin work. In contrast, consider the best planning process suggested by Keles, with open collaboration, debates on the validity of assumptions, and frequent model updates. We describe this as a **short cycle**—it may well describe the patterns of work that occur *within* each functional team among collegial experts in the same domain.

Is it possible to achieve greater productivity by transforming the long cycle into a short cycle? In some cases, we believe the answer is yes. One sweep through the long cycle is probably required to collect data and establish a baseline for planning scenarios. But if management determines that adjustments to the baseline plan are needed, *the GridIntell framework enables collaboration in the manner of the short cycle to make the adjustments quickly*—thus avoiding multiple traversals of the long cycle.

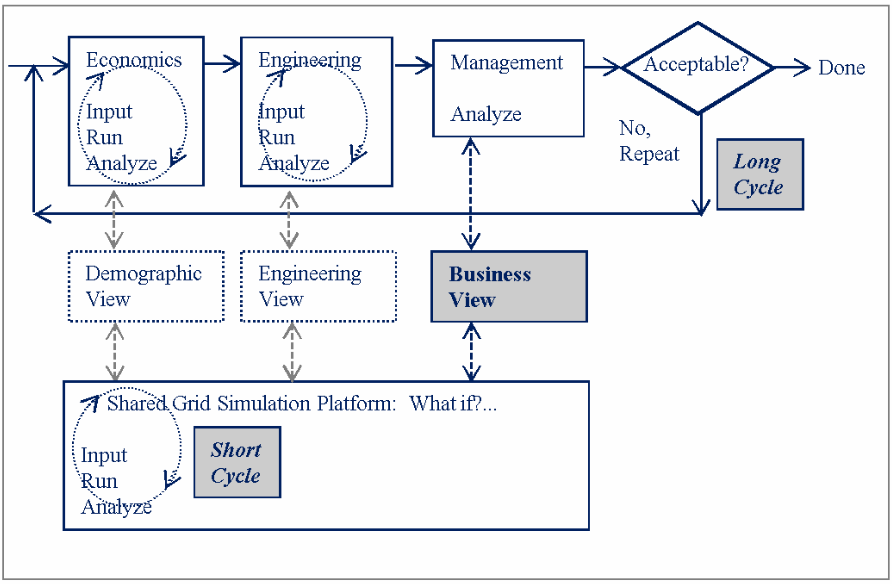


Fig. 1. The process of building accurate grid simulations involves multiple kinds of functional expertise. Here is a simplified example showing three types of contributors: economists, engineers, and managerial decision makers. The serial path through all functional domains can take significant time, which becomes a problem when the entire process must be repeated to answer new questions or fix faulty assumptions. The GridIntell framework allows some kinds of repeated analysis to move from the long cycle to the short cycle.

3 Implementing the GridIntell Framework

GridIntell provides two process elements to enable short cycle-style planning. The first is a set of graphical user interfaces shown in Fig. 1 as the Demographic View for economists, the Engineering View for engineers, and the Business View for managers. The second is the Shared Grid Simulation Platform, that can rerun scenarios as desired from any of the GUI views.

Each view offers a different set of parameters for a user to vary, and reports different key performance indicators (KPIs) to the user once a simulation is run. Table 2 gives some examples. The engineering team for instance seeks to determine if the quality of electric service will be reduced due to overutilization, and observes values for KPIs such as the Cumulative Average Interruption Duration Index (CAIDI) for simulated neighborhoods. Managers focus more on financial KPIs. The impacts of the economists' data are measured in the Engineering and Business Views, so economists can access KPIs from those views.

Table 2. Examples of how the three different views interact with the Shared Grid Simulation Platform of Figure 1

View	Modeling Topic	KPI Output Example
Demographic	Electric vehicle adoption rate	--
Demographic	Population and business growth	--
Engineering	Utilization of grid assets	Transformer utilization rate
Engineering	Stress on grid assets, blackouts	CAIDI
Business	Capital expenditures	CAPEX per kWh
Business	Revenue from operations	Revenue per kWh

In fact, we limit the scope of each view mainly to reduce complexity for the user, and not prevent access to results—we include menu options in each view to access all KPIs across all views. Depending on how the workflow is defined, we may not share privileges for uploading all types of data into the shared simulator. For instance, only users permitted to login to the Engineering View should be allowed to upload power engineering models into the Shared Grid Simulation Platform.

Fig. 2 shows how the four major software components of the Shared Grid Simulation Platform work together to simulate a model. A user logged into any one of the three GUI views may run simulations, though again the parameters they are allowed to vary will depend on the view. Once the run command is given, an On Demand Scenario Builder compiles a simulation source file. Data for this file may come from critical enterprise databases—customer relationship management (CRM), geographic information systems (GIS), enterprise asset management (EAM), weather databases, and predesigned scenario templates that define how to instantiate entities within simulations.

Note that these templates in turn must be designed by experts, and that design work is part of the long baseline planning cycle. The GridIntell framework envisions template-building tools in the Demographic View and Engineering View, but we will defer the discussion of those aspects to future papers.

After the source file is compiled, the Scenario Builder submits it to the Simulation Executive to be run. When the run finishes, business intelligence (BI) analysis tools format raw output data, run additional statistical algorithms, and present visualizations of KPI trends to the GUI, completing the short cycle.

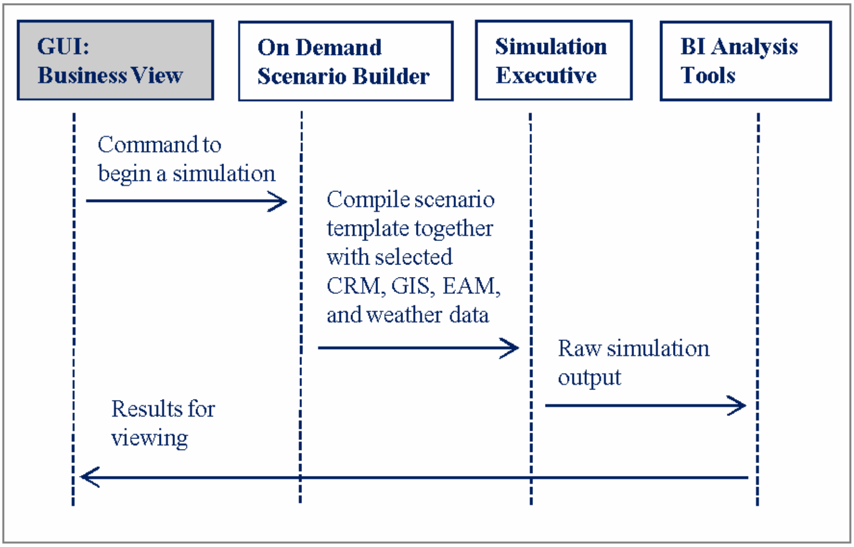


Fig. 2. The GridIntell solution framework contains four types of software. The user interacts with a graphical user interface—here we chose the Business View from Figure 1. After entering necessary configuration settings, the user hits "run", and the succeeding elements execute in turn to provide simulation results.

4 GridLAB-D as the Simulation Executive

At this point we have defined the GridIntell Framework such that users with different types of expertise may collaborate on a shared model—and this framework may conceptually achieve the value proposition of better and faster forecasts. It should be noted that three of the four categories of components in Fig. 2 have common implementations. For example, the GUI may be a browser application; the On Demand Scenario Builder could be a program written in a scientific scripting language; and there are many options for the BI Analysis Tool, from spreadsheets to more powerful enterprise software.

However, the Simulation Executive component is unique—it must be comprehensive enough to run models of interest to users of all three views. In our research we determined that an open source power system simulator from Pacific Northwest National Labs named GridLAB-D has the requisite functionality to fulfill this role (see e.g. [3],[7],[8],[9]). Table 3 lists some of its features.

For use within GridIntell, two characteristics of GridLAB-D stand out: it runs discrete-event models, essential to some types of economic modeling; and it computes the physical state of the grid using power flow solvers, which is essential to engineering calculations. It is not used primarily as a statistical calculator for methods such as regression analysis, but such features are commonly available in BI analysis tools, and so the fourth component of Fig. 2 can take care of those other modeling requests.

To illustrate how electric power utilities may use the simulation power provided by GridLAB-D to develop forecasts, we posit a hypothetical full-service California utility company called CalPower. The CEO of CalPower delegates to a planning team the mission of building forecasts of necessary generation capacity and multi-year investment plans. The team members are modeled after the user roles described in the Introduction.

Table 3. Selected properties of the GridLAB-D power system simulator

Property	Specification
Type	Agent-based discrete event simulator
Economic models	Retail power market models
Engineering models	Choice of power flow solvers, climate models, library of generation, transmission, and distribution objects
Connectivity features	Weather files, event definition files, connectivity to common databases

We consider a scenario based on California Executive Order S-21-09, which defines a Renewable Power Standard (RPS) mandating that by 2020 at least 33% of all electrical power delivered comes from solar, wind, geothermal, biomass or small-scale hydroelectric sources [10]. CalPower currently has achieved a 15% RPS. With the 33% RPS goal—as well as other goals such as reducing carbon dioxide emissions and meeting extra power demand from new electrical vehicles—the team at CalPower begins a new planning cycle to investigate solutions. The team uses the GridIntell framework to simulate the changes that will happen to the power grid between year 2011 and year 2020.

Through GridIntell, the planning team has uploaded scenario data such as CalPower's current generation portfolio, infrastructure and load models, market trend forecasts, and weather data. They also enter investment plan options that they can later evaluate according to KPI forecasting results provided by the simulation engine. Users with different expertise collaborate on these inputs within the GridIntell Framework. For example, the economists contribute market trend data and investment strategies, while the power engineers model the infrastructure and load at high resolution. Based on this collaborative work, the team built three investment strategies to compare: Plan A and B respectively emphasize wind and solar, while Plan C is more balanced between wind, solar, and biomass technologies.

With these input data, GridIntell generates KPI forecast values that describe the benefits and risks of each investment plan. The planning team will be able to summarize the findings similar to Table 4. Note that each plan has its own advantages and disadvantages due to its technology portfolio. For example, wind and solar are more unpredictable than biomass generation, but do not produce carbon dioxide ([11], [12]).

The simulation results not only help the planning team compare investment strategies, but also allow deep dive into the forecasting numbers to gain insights into the future. For example in Fig. 3, an analyst user drills down into a certain simulated year to observe the daily renewable energy generation pattern.

Table 4. The GridIntell simulation result allows the user to compare KPIs across different investment plans. Plan A and B are subject to high exception risks because wind and solar technologies are volatile and difficult to predict. Plan C is more steady thanks to the balanced portfolio, but is "dirtier" than Plan A and B because of its higher usage of biomass technology.

Plan	Advantage	Disadvantage
A, wind focus	Lowest total cost	Lower service reliability
B, solar focus	Lowest OPEX	Lower service reliability
C, biomass focus	Lowest CAPEX, highest reliability	Highest $CO2$ emissions

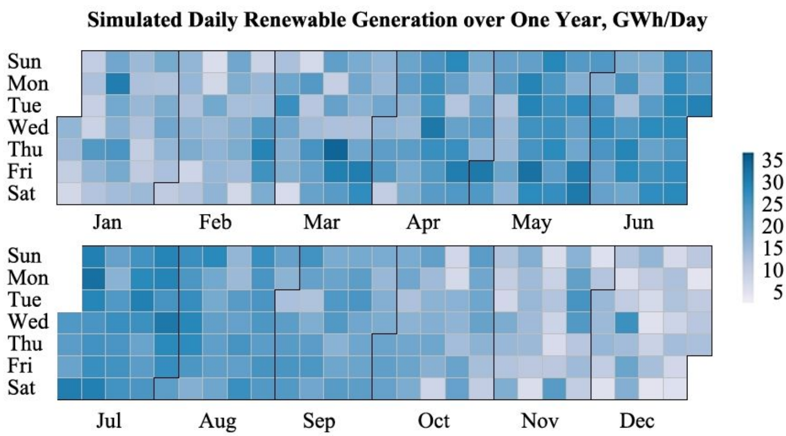


Fig. 3. As part of the RPS planning scenario, the user can obtain visualizations of renewable power generation totals over time. Here we show renewable energy production and use each day over the course of a simulated year. Two properties of the data stand out: there is considerable variation in the power output from day to day, and peak renewable generation in this region occurs in the summer. In the GridIntell framework, GUI users may also drill down by clicking on a day and observing higher time resolution trends that affect each diurnal power cycle.

5 GridIntell User Experience Validation Research Plan

The GridIntell Framework is currently a research project, and so is considered to be in the *conceptual phase* from the perspective of our user experience designers [13]. Based on our preliminary research, we have developed as phase two of the project a robust user experience validation plan that ensures GridIntell will meet the needs of utility planning teams. Key topics from this plan are shown in Table 5.

The conceptual phase provides an opportune time to organize information architecture and gauge future usability issues. In parallel with the user buy-in stage we just completed, user research obtained customer participation in future phase two user experience methods. The customer buy-in for both the solution and acting as customer innovation partners has provided us the foundation for phase two research.

Table 5. Areas for future User Experience (UX) validation research on GridIntell

Research Topic
1. Build detailed models of the critical work flows for each role in Table 1.
2. Enumerate reasons that planning teams repeat the long cycle (Fig. 1).
3. Determine what artifacts are used to justify decisions.
4. Further explore how GridLAB-D functionality complements existing tools.
5. Evaluate the confidence of planners in existing forecasting methods.
6. Validate that the GridIntell framework improves timeliness and accuracy of forecasts.

Our research plan is informed by recent works in the collaborative design space—from lessons learned, to methodology for design. Hupfer et al. have developed a series of collaborative design guidelines to aid in the development of solutions with a collaborative component [14]. Some approaches we use include:

- *In-depth interviews*, a semi-structured open-ended interview with questions centered on specific themes: tools, work flows, collaboration points. Audio and text transcriptions are gathered to develop user profiles.
- *Site visits*, to observe users collaborating in familiar surroundings, and convene focus groups of users [15]. Notes from these visits help us build affinity diagrams that inform UX scenario design tasks.
- *Cognitive walkthroughs* to validate initial low-fidelity mockups from the user interface design team [16].

Bedekar and Kennedy [17] illustrate the importance of user research for achieving sustainability goals, and show that an important opportunity for energy management solutions lies in revealing unknown processes to lay a foundation for new and innovation collaborations. Our user research plan seeks to reveal these unknowns and allow us to fulfill GridIntell's value proposition.

6 Conclusion

In this paper, we propose the GridIntell Framework with three key attributes that support improved capital investment planning forecasts for electric power utilities. First, it provides a collaborative workbench for efficient cooperation between experts in different domains that are vital to the planning process. Second, it provides comprehensive grid performance modeling and forecasting capabilities with the integration of GridLAB-D, a powerful discrete event simulation engine. Third, it offers a superior user experience with easy-to-use yet powerful visual analytics that reduce learning curves, improve individual productivity, and drive valuable business insights. We propose to build on our research and prototype development here with new user experience validation studies based on sound principles of collaborative design research.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant # EEC-0946373 to the American Society for Engineering Education.

References

1. Climate Change—At Least Some States and Cities Get It. New York Times, January 9 (2011)
2. Blumsack, S., Samaras, C., Hines, P.: Long-Term System Investments to Support Plug-In Hybrid Electric Vehicles. In: Power and Energy Society Meeting: Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, Pennsylvania, USA (2008)
3. Ortmeyer, T., Dugan, R., Crudele, D., Key, T., Barker, P.: Renewable Systems Interconnection Study: Utility Models, Analysis, and Simulation Tools. Technical Report, Sandia National Laboratories (2008)
4. Podmore, R., Robinson, M.: The Role of Simulators for Smart Grid Development. IEEE Trans. Smart Grid 1(2), 205–212 (2010)
5. Keles, D., Most, D., Fichtner, W.: The Development of the German Energy Market Until 2030—A Critical Survey of Selected Scenarios. Energy Policy 39, 812–825 (2011)
6. Kulshrestha, P., Wang, L., Chow, M., Lukic, S.: Intelligent Energy Management System Simulator for PHEVs at Municipal Parking Deck in a Smart Grid Environment. In: IEEE Power and Energy Society General Meeting, Calgary, Alberta, Canada (2009)
7. GridLAB-D, from Pacific Northwest National Labs, http://www.gridlabd.org
8. Aliprantis, D., Tesfatsion, L., Zhao, H.: An Agent Based Testbed for the Integrative Study of Retail and Wholesale Power System Operations. Working Paper, Iowa State University, http://www2.econ.iastate.edu/tesfatsi/RetailWholesalePaper.ATES2010.pdf
9. Fuller, J., Schneider, K.: Modeling Wind Turbines in the GridLAB-D Software Environment. J. Undergraduate Research, U.S. Dept. of Energy 9, 66–72 (2009)
10. California Executive Order S-21-09, http://www.cpuc.ca.gov/PUC/energy/Renewables/hot/33implementation.htm
11. Castaner, L., Silvestre, S.: Modeling Photovoltaic Systems Using PSpice. John Wiley and Sons, West Sussex (2002)
12. Piwko, R., Osborn, D., Gramlich, R., Jordan, G., Hawkins, D., Porter, K.: Wind Energy Delivery Issues: Transmission Planning and Competitive Electricity Market Operation. IEEE Power and Energy Magazine 3(6), 47–56 (2005)
13. Start Here: The Wheel Model for User Experience, http://www.uxexpert.com/the-wheel-model-for-user-experience/
14. Hupfer, S., Cheng, L., Ross, S., Patterson, J.: Introducing Collaboration into an Application Development Environment. In: ACM Conference on Computer Supported Cooperative Work, Chicago, Illinois (2004)
15. Twidale, M., Nichols, D.: Designing Interfaces to Support Collaboration in Information Retrieval. Computers 10, 177–193 (1998)
16. Using Prototypes, http://www.sapdesignguild.org/community/design/prototypes.asp
17. Bedekar, N., Kennedy, S.: What Sustainability Brings to the Research Table. In: Proceedings of the 22nd Annual Conference on Human-Computer Interaction, Orlando, Florida, USA (2011) (forthcoming)

The Application of the Concept of Affordance to a Creative Design Method

Chien-Kuo Teng[1,2] and Ming-Chuen Chuang[2]

[1] Department of Industrial Design, Shih Chien University, 70, Da-Zhi St. 104 Taipei, Taiwan
[2] Institute of Applied Arts, National Chiao Tung University, 1001 University Road, 300 Hsinchu
teng@mail.usc.edu.tw, cming@cc.nctu.edu.tw

Abstract. This research integrated ideas regarding affordance into a method for creative design, including five major steps: 1) Observe behavior; 2) Note down events happening or issues; 3) Figure out pattern; 4) Obtain messages of behavior perception; and 5) Reinforce message and naming. In this research, an "*Another Hand*" plate was designed to demonstrate the applicability of this method. In the application process, first, the researchers observe the dining environment, and discovered that it was difficult for people to scoop up the last few bits of food on the plate; therefore, they need the aid of extra tableware to finish their food. Then, observing users, we discovered that the spoon might turn around the plate. Next, we transform the idea into the design of the wedge to stop the food from moving. So the diner can finish the last bits of food using a natural eating motion. Lastly, with "*Another Hand*" plate as an example, we observed the process of 10 people eating with this plate. It was discovered that users could use this wedge to help scoop the food at 64.2% of the dining time. Hence, this invention won the final list design award of I.D.E.A. held in the USA in 2008. This research intends to review and revise this design method for better design for the reference of other inventors.

Keywords: affordance, observation, arousal, experience, design method.

1 Introduction

Product availability and usability are the primary key points during the process of industrial design. Via the product, it is necessary for designers to make users aware of appropriate user direction of the product and what the product can accomplish. Norman (1998) has pointed out that useful design is the criteria for future product competition. Over recent years, product designing has been emphasizing more on user experience and instinctive operation. Myerson (2001) stated that a new age of design, focusing on the whole experience of user interaction, required more subtle calibration. The development of psychology will also be studying and interpreting humans' subconscious behavior from another perspective, providing the product design more humanistic consideration in regards to the interaction between humans and objects. Of which, the concept of affordance evolved from ecopsychology has further expanded its view on humans' behavior to be the result as developed from the mutual influence

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 215–224, 2011.

between living things and environment, which is a perspective different from the concept of message handling as emphasized in the traditional cognitive psychology (Gibson, 1979; Mark, 1987; You, et al., 2007; Goto, 2008). Affordance is a direct perception theory created by eco-psychologist J.J. Gibson, which means the possibility that the environment endows upon a living thing's behavior. The concept itself is not a designing theory or method, but many designers today have extended it to application in the field of design. In the past, there has been study (Huang Rong-cun, 1991) believing that Gibson's contribution is on the conceptual separation of imposed and obtained stimulation. Donald Norman (1998) used daily necessities to explain ways to strongly hint at an object's operation via affordance. For instance, a flat surface could be pushed, a round button could be turned, or a hole could be inserted. It could help designers during their product designing stage to enhance the products' availability and usability as they are being operated by their users, which would then allow users to more instinctively operate the products.

Attention to user's behavior during product operation has gradually become a designing trend around the world. Creation from product designer Naoto Fukasawa and the merchandise at MUJI have been widely accepted and enjoyed by many. Each believe that they have brought forth, such as "Without thought", "Later wow" and "What's happened naturally", has carried the following messages: user would ultimately understand product's design value via the behavior of implicit product operation; and the interaction between product and user could be made more natural. This point of view allows designers to have a more objective source of judgment when they ponder the interactive relationship between product and user as well as a perspective of wider vision and higher diversity. Thus, this research made the hypothesis on whether there could be a more concrete way or designing pattern, which can provide a foundation for designers to observe, step-by-step, the affordance that can be provided by the environment and then transform this observing experience in the proposal of an actionable designing result? Based on this goal, we are summarizing possible applicable methods for this affordance below, and we are also proposing a designing pattern and then applying the pattern in the design of a case study. We are hoping to build a preliminary design pattern that can stimulate design imagination based on practical design.

2 Literature Review

Zeisel (1984) believed that designing, an intricate activity, connects three basic activities: imagination, presentation and testing. And there are two categories of knowledge needed during designing according to their purpose: 1) knowledge that initiates a catalyst during imagination; and 2) knowledge that serves as test. (Korobkin, 1976; Zeisel, 1984) pointed out that parts of the designing knowledge are able to be observed via physical environment or environmental behavior. In regards to actual designing practice, designers would need to possess a sharp sense of touch towards the environment. IDEO founder Bill Moggridge proposed five designing steps: 1) understand, 2) observe, 3) visualize, 4) refine and 5) implement (Takeshi Goto, Masato Sasaki & Naoto Fukasawa, 2008). Among these steps, understand and observe were identified as the top two elements in the designing process. Thus, the importance of these two steps is apparent.

Affordance is the main concept under the direct perception theory of ecopsychology. Transforming the verb, afford, to a newly invented noun, affordance, James J. Gibson created this word to describe the specific relationship between living things and their surroundings (Gibson, 1979). Gibson believed that the messages that are influencing our perception and activity have already existed in the external world. And whether the observers could pick up these useful invariants would depend on observers' "purposeful" and proactive pursuit. The study explained the role "proactive discovery" plays in the process of searching for the invariants of form perception. Affordance has also been considered as one of the important designing factors in the enhancement of product usability and interactive capability (You &Chen, 2007). In regards to the interpretation of "affordance", it has been interpreted with a following statement - "When the affordance of an object or the environment corresponds to the functions as expected, the design would exhibit with higher efficiency and easier use" (Lidwell, 2008).

Designers would also sometimes be accustomed to the existing method and thus be presenting ideas via the stereotypical or routine format. And they would thereby forget to observe the operation behavior that is actually taking place with a different attitude and find other potential opportunities to show off their design. The designer's subjective cognition, change of the environment's movement, and user's the behavior pattern would all influence the user's interpretation of the product language. Therefore, designers would need to pay attention to and consider the context within which users are using the products. This concept has been inspiring for the contemporary designers who have been seeking change. Designers are able to obtain a behavior's potential possibilities through observation. Via a designers' interpretation, they are able to rethink ways to operate instinctually and allow users to understand the operation via a form that is instinctive to the general public. As new vocabularies that are compatible with behavior are interpreted on the most apparent aspect of a product, the playfulness and the meaning of designing pieces would then be enhanced. Consequently, the comprehension gap between designer and user would also be reduced. This kind of proactive designing method is exactly the so-called perceptive message of reinforced behavior. Lastly, "design naming" would be used as a strategy to stir up user attention.

3 Observation and Creation Methodology

This research is planning to integrate the affordance concept as proposed in the field of psychology in the division of the environmental messages' observation methods. Moreover, we are also planning to discuss the interpretation method of affordance from a perspective of real case creation. We are trying to induce a more concrete creation principle and then proceed onto design creation with this methodology, so that we can provide designers a reference principle while they attempt to execute designing with the application of affordance. After going through literature organization and introduction of Chapter Two, we have obtained the three major characteristics of affordance that correspond to product design: 1) Utilize observation of the environment; 2) Activity-Centered Behavior Method; and 3) Point out perceptive messages symbolizing affordance. Then the three characteristics are further divided into five steps: 1) Observe behavior; 2) Note down events happening or issues; 3) Figure out pattern; 4) Obtain messages of behavior perception; and 5) Reinforce message and naming, so that we can further propose actual execution methods during the designing procedure. We have derived a preliminary procedure that applies affordance, which is displayed in Figure 1. We are explaining below in detail the process of applied affordance creation methodology as proposed in this research.

Utilize Observation of the Environment		Activity-Centered Behavior Method	Point out Perceptive Messages Symbolizing Affordance	
1. Observe Behavior →	**2. Note Down Events Happening or Issues →**	**3. Figure out Pattern →**	**4. Obtain Messages of Behavior Perception →**	**5. Reinforce Message and Naming**
Observe the environment for behaviors that have already happened, the space or the byproducts being used.	**Examine ongoing behavior.** Then break up the motion during the activity. ←	**Record the physical characteristics of the material.** Exhibit product uniqueness of the new design. **←Examine issues**	**Obtain attributes and apply:** behavior comparison and designing method to proceed onto designing. **←Behavior comparison**	**Name after examining** performance result or user behavior after the change.

Fig. 1. The method of five steps and its flow chart to applying the concept of affordance

(1) Observe behavior: Designers need to observe the behavior in the environment that has already happened in the "past". Observation of the behavior that has already happened, space or byproducts that have been used include deterioration, remains or lack of signs (Zeisel, 1984). Deterioration refers to the wear resulted from frequent use, e.g. a small trail formed as people walk over the grass, or the slits left behind by knifes on a meat cutting board. Remains refer to things left behind from certain activities, e.g. the empty wine bottles after a banquet, cigarette ashes in the ash tray, opened food cans or scattered chairs. Through observation of the environment, events happening or issues are found. And no personal judgment shall be added during the observation so to avoid the influence from stereotypical image.

(2) Note down events happening or issues: Since activities are considered as a sequential process, videotaping or follow-up can be used to describe a complete activity procedure. Examine the behavior that is happening at the moment, including the relationship between objects and people as well as the relationship between objects and objects. Later, take apart the motions within the activity and examine further this "user habit".

(3) Figure out patterns: Designers can record the physical characteristics of a material, including shape, structure, hardness and surface pattern. Secondly, refer to people's existing user habits, e.g. trembling their feet while they ponder, unconscious hand motions while they dine, coordination of two hands' motion, etc. The information will be the base for the entry point where users begin their relationship with products so that the new design's product characteristics would be exhibited.

(4) Obtain messages of behavior perception: Obtain and apply observed signs of related behavior messages. The first application would be a comparison of similar behaviors. And the second application would be to accentuate or exaggerate the meaning displayed via a designing method (e.g. repetitive motions or objects that replace this motion).

(5) Reinforce message and naming: Through designers' naming per performance result or by describing the user behavior after transformation, designers can reflect on if the design is able to communicate the message they were hoping to pass on. Second, designers should consider the reminding effect that naming can place upon users. When users are using the product under specific circumstance, appropriate naming can help users to recall that special memory or experience of theirs. Thus the product can be made more easily accepted.

4 Design Creation

Taiwanese people eat mainly rice dishes, and fried rice is one of the most representative dishes in Taiwan within its rice eating culture. Fried rice is usually served in a shallow plate with wide mouth. We use it frequently, but we seldom notice where it needs to be improved. This design creation has been done via a planned designing pattern with "dining behavior design" as the theme (designer: Ying-Ting Kuo). During observation and actual practice of the procedure, we also discover different possibilities hidden within. We organize the designing steps and present the key points as below.

4.1 Observe Behavior

Designers noticed that there was often food left on the dining plate and thus found out this issue. Whether people were dining with chopsticks or spoons, or having food that was served on a plate, it was always difficult to finish that last bits of food. And when using other utensils to supplement the dining process, its purpose was exactly to assist the behavior motion of "stopping". A corresponding relationship exists between spoon/dining plate and people.

4.2 Note Down Events Happening or Issues

When we continued observing the actual dining process, we noticed that people would be accustomed to using another hand to hold other utensils (e.g. another spoon) to supplement their dining and finish the last bits of food. In terms of the reasons that make it difficult to scoop up the last bits of food, it could be that the food in the plate was reduced, the weight was also decreased. Correspondently, affordance that could stop the spoon to scoop up the food was also reduced. As a result, the food would be continuously pushed to the rim of the plate. We set a user behavior for our design according to the issues discovered at this stage: Don't overly change the original dining habit and allow users to finish their food from the plate more easily. Or in other words: Keep the form of the round-shaped ceramic dining plate, and find ways to increase the probability of finishing up the food. Sometimes there would be noise made when the spoon touches the rim of the ceramic plate or another supplemental utensil, which is also an indication for the users.

4.3 Figure Out Pattern

Possible motions that could succeed in supplementing the goal of scooping up the food include: stopping, transferring and halting. According to the dining behavior observed above and the pre-set goal, we chose “stopping” to be our preliminary designing concept. Since the round-shaped dining plate originally possesses a shallow, thin and flat shape that is meant to hold and display food, it is not suitable to make any changes to its flat part. Also, because that last bits of food could not be stopped, they would be pushed to the rim of the plate. As a result, we had reason to think about our design starting from the plate rim. On the rim of the round-shaped dining plate, we designed a vertical convex stopping wall. When users found it difficult to scoop up the last bits of food, they would naturally push the food towards the convex stopping wall, which would help them to scoop up the food. Step 3 of our design creation ends here, and it has already provided a basic behavior possibility for the anticipated user behavior. In Step 4 that follows, we will be thinking about ways to reinforce the perceptive message or behavior image through further styling method. Besides allowing the anticipated operation to be established, the following step will also ensure our designing product has additional meaning.

4.4 Obtain Messages of Behavior Perception

In this step, designers started to think about related messages and signs of behavior that could further transform onto product’s style and guide the users. In Step 1, it was observed that people have the habit of using another hand to supplement their dining process. This is a behavior that people have naturally developed in the environment. And following this universal behavior, signs of this behavior can serve as a clue for further styling transformation. Possible forms of signs include a hand and spoon, or the image of other utensils. Under the prerequisite of Step 2, we chose the image of a curved palm (Figure 2) to be the transformation of behavioral signs. With a fading surface, an invisible hand-like design is extended out along the rim of the round-shaped plate to help users during their dining process. The design corresponds to the events observed in Step 1 and places upon the product operation a more significant meaning. Thereby a new perceptive message was born.

4.5 Reinforce Message and Naming – *Another Hand*

Naming this design *Another Hand* corresponds to where the idea of this design comes from. We are hoping that users would realize the meaning of this image we have designed and the advantage of normal people having another hand to assist them during dining when they are near the end of their meal and have just discovered the behavior possibility hidden (using the convex stopping wall to stop the food).

4.6 Design Result: Presentation of the Uniqueness of *Another Hand*

The spotless white round-shaped ceramic dining plate has always been the best complement that brings out the arrangement of the food. However, the shallow and thin shape has often made it difficult for diners to finish every bit of the food when there are last bits of them left. Under the condition of not affecting the round-shaped dining plate, this design created a vertical wall to assist diners as they finish up last bits of food with our styling method. Even when the amount of food reduces, users can still push the portion of food left to the convex stopping rim. Through transforming the image of a curving palm to be the convex shape and naming it as *Another Hand*, it is as if an invisible hand is helping people to compete their meal. At the same time, it is also reminding us that we shall not waste a single bit of food. Considering user habits, the design keeps 70% of the opening arc area to allow both the left and right hands to operate smoothly. The angle at the convex stopping rim is easy for demolding manufacturing so that plates can still be piled up for storage. Figure 3 summarizes the designing thinking procedure described above.

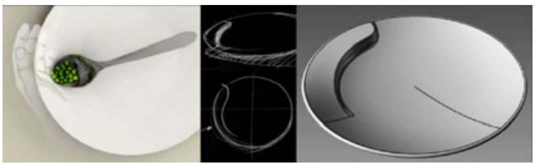


Fig. 2. Transformed image of hands to a concrete designing project.

Fig. 3. *Another Hand* with opening arc area to allow both the left and right hands to operate smoothly, can piled up for storage.

5 Design Result Assessment

In regards to the creation result as brought forth by this research, we have performed an assessment via the two aspects of expert interview and user survey.

5.1 Expert Interview Assessment

In order to obtain more objective performance assessment for this research's creation result, this research invited two experts from the field of psychology and designing to provide our creation with some suggestions in the form of interviews. The expert from the field of psychology was a professor of Design Psychology at a college's dep. of industrial design. The expert from the field of designing was a senior designer with 12 years experience. The interview was separated into three parts: see, listen and interact. First, the expert individually observed the product without receiving any user instruction. Then, the designer described the circumstance and object for which the product was designed to. Creation though process was also explained. Lastly, interaction with experts was held for overall discussion. Organized below are the key comments that these two experts provided for this research's creation:

(1) Psychology Expert A: The work *Another Hand* provides a certain level of effective and appropriate affordance in terms of the food-stopping behavior. However, different results happen when encountering different food and utensils under the influence of different eating cultures. For example, when using a fork while having pasta, this stopping function might not be necessary.

(2) Senior Designer B: This work chose the method of a vertical stopping surface to provide the food-stopping function, when the user is holding up the plate, the flat surface area at the rim of the vertical stopping surface can also provide the behavior possibility for the fingers to be placed on top of it. This will enhance the stability when the user holds up the plate and avoid the fingers from touching thee food served inside the plate.

5.2 User Survey

This research performed a user survey via the experimental method, where we observed how each subject finished the same portion of fried rice using spoon. There were a total of 10 subjects; each time the subject used the spoon to scoop up a spoonful of fried rice, a picture was taken. Following this procedure, we recorded down each user's behavior reaction and feedback towards the convex stopping part of *Another Hand* while they were having the fried rice. Since the speed of their dining process was not the focus of this research and to avoid any influence that timing might have on diner's habit, this experiment did not record the time it took for individual subject to finish eating. Each subject scooped up different amount of fried rice in each scoop. We defined 100% as the total number of scoops each individual took to complete the dining process from their first to their final scoop. Divided the total number of scoops by the number of scoops of fried rice picked up with the help of the convex stopping part, we could derive the progression ratio of the advantage of "used" convex stopping. Per the experiment result, we found that among the number of spoons it took for the 10 subjects to finish the fried rice, the minimum was 7 spoons and the maximum was 12 spoons. Some people made use of the convex stopping wall, as soon as 10%, and some discovered the function as late as 90%. And concluding the fried rice eating process for all 10 subjects, the overall average progression rate of discovering the function of and using the convex stopping part to assist in scooping up the fried rice was 64.2%. Survey results of *Another Hand* users are summarized in Table 1 below.

Table 1. Summary of *Another Hand* Dining Plate Usage Survey Result

Item \ Subject	1	2	3	4	5	6	7	8	9	10	Average
Total Spoon Count	10	10	10	7	12	10	11	8	10	8	
First Use of Convex Stopping Wall	9	1	8	6	9	6	7	5	4	6	
Progression Ratio	90%	10%	80%	85.7%	75%	60%	63.6%	62.5%	40%	75%	64.2%

Overall, there was no apparent difference between the performance of regular users and that of users with design-related background. We also discovered during the process that there were users who tilted the plate to increase the effectiveness of scooping up the fried rice. And there were also people who could use their left hand to use the design without any issues. Besides the scooping motion, we also observed users' unconscious motion of dangling their spoons while they chewed. Besides the motion of using the spoon to scoop up, there were also incidents where users were reminded of the function of convex rim as their spoon made a sound while it circled around the plate rim. The abovementioned observations all made it apparent that users were able to understand the hidden function of *Another Hand*.

5.3 Brief Conclusion of Designing Results

In terms of designing results, the work *Another Hand* participated in Taiwan's 1st Annual Universal Design Award in 2006 and received an Excellence Award. After enhancement was made on the design in 2007, i-use was authorized for its mass production, manufacturing and sales. In 2008, the work received the 2008 I.D.E.A. – International Design Excellence Award of the U.S. A possible reason for the recognition of the work *Another Hand* lies in its fine integration between its functionality and the style symbolism of the behavior exhibited, which at the same time maintains the plate's original function diversity before the innovative design. Results of the user survey also showed that users might not notice the behavior possibility of the convex stopping part when they first started having the fried rice, but when the amount of fried rice was reducing, they would naturally push the fried rice towards the convex stopping part. And this phenomenon also corresponds to what the creator has originally anticipated. On the other hand, on the aspect of the development process of this research, it began from building a preliminary design of thought pattern to actual product manufacturing. And there were gradual adjustments on the organization and the writing of concept details as well as product details and pattern, which spanned a total of four years of development and revision. It has been apparent that the designing pattern is able to assist designer's thought process and possesses a preliminary application value.

6 Conclusion

This research proposed the three procedures of 1) observation of the environment; 2) activity-center behavior; and 3) connection with behavior image, along with five derived steps. Design case studies and evaluation were also used in the advanced discussion of the creation result. Even though its form of display was to bring the dining plate design to correspond to user's dining behavior and to enhance the convenience of the dining plate, there were still subjective differences during observation of the environment. After proactively figuring out the message, transformation was still needed to show off the existing affordance performance as a designed object. One thing worth mentioned is that the message mentioned here does not only direct to the consensus resulted from users "looking onto" the product, but it can also be viewed from a more expanded perspective to be a reflection of behavior, which is something that needs to be brought forth after a process of body contact. For the users, the message relates to that corresponds to affordance in the original environment after final design naming via the process of product operation. We are hoping future designers may discover more designing opportunities through the observation methodology described in this research. Listed below are four major conclusions and suggestions we have summarized from this research.

1. In terms of the structure of designing pattern, affordance is originally a theory proposed by eco-psychology. Through discussion of related theories and documentaries as well as designers' creation experience, this research came to realize that "behavior" could be the main entry point from where it could find out a product design that corresponds to affordance. For the design creator, possessing a behavior-orientated integrated advantage would assist designers during their designing process to create a style that would pass on the behavior message. Starting from planning product operation for the search of usable clues of affordance during the behavior procedure, product's usability, exterior design and reflective meaning would all be integrated step by step.

2. With regards to "The Design Procedure of the Applied Affordance Concept" as proposed by this search, even though it benefits designers in showing the structure of behavior message during their designing procedure, they still need to rely on their individual experience and aesthetics judgment during actual styling. Designing result can make the affordance design exceedingly explicit, and it can also make it utterly implicit. Overly exaggerated or poorly handled product messages can also create inappropriate association for the users that may cause errors during their operation.

3. Dining a plate design needs to take into consideration the size of the grains, the humidity and stickiness of the content (e.g. fried rice). These variables all influence the friction of the dining plate and change the method used to pick up the content inside the plate, and certainly they also influence the style of the product. Follow-up research can be based on the consideration abovementioned to integrate holding capability and unexpected effect as it looks for other applications.

References

1. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin Company, Boston (1979)
2. Goto, T., Sasaki, M., Fukasawa, N.: The ecological approach to design. Azoth Books, Taipei (2008)
3. Huang, R.C.: The Psychology and Behavioral Researcher (1991) ISBN 957-09-0482-8
4. Krippendorff, K.: Design: A Discourse on Meaning. A Work Book. University of the Arts, Philadelphia (1994)
5. Lidwell, W., Holden, K., Butler, J.: Universal Principles of Design. UN Books, Taipei (2008)
6. Myerson, J.: IDEO: Master of Innovation. Laurence King Publishing, London (2001)
7. Norman, D.: Human-centered design considered harmful. ACM Digital Library Magazine (2005)
8. You, H.-c., Chen, K.-S.: Applications of affordance and semantics in product design. Design Studies 28 (2007)
9. Zeisel, J.: Inquiry by Design-Tools for Environment-Behavior Research. Cambridge University Press, Cambridge (1984)

A Product Design Approach by Integrating Axiomatic Design and TRIZ

Shiaw-Tsyr Uang[1,*], Cheng-Li Liu[2], and Mali Chang[1]

[1] Department of Industrial Engineering and Management, Minghsin University of Science and Technology, Hsinchu, Taiwan
[2] Department of Management and Information Technology, Vanung University, Taoyuan, Taiwan
uang@must.edu.tw

Abstract. The purpose of this research intends to integrate the strengths of axiomatic design (AD) theory and theory of inventive problem solving (TRIZ). This study establishes a systematic product design model by adopting some major tools from AD and TRIZ such as functional requirements, design parameters, design matrix, contradiction matrix and inventive principles. Furthermore, the proposed model's efficiency is analyzed and evaluated by a case study of a Handheld GPS product. Results indicate that the design model which combines with two theories can find out the usability problems and solutions efficiently. When applying the proposed model on product redesign or new product development may avoid the cost waste and increase the design efficiency and usability during the product design and development processes.

Keywords: Theory of Inventive Problem Solving, TRIZ Theory, Axiomatic Design, AD, Human-Machine Interface Design, Product Development, Hand-held satellite omniselector.

1 Introduction

Nowadays, there are different kinds of products in the consuming electronic markets, and the competitions among these industries are severe fiercely. Thus, in order to gain a significant market share in the world, products have to be not only innovated efficiency but also fitting the consumer requirements effectively.

Axiomatic design (AD), developed by Nam Pyo Suh, is a human-machine interface design tool using matrix methods to systematically analyze and transform customer needs into functional requirements (FRs) and design parameters (DPs) [2]. The relationship between functional requirements and design parameters is represented in a design matrix. Good (decoupled) designs can be represented by nxn triangular matrices, e.g., all entries above the main diagonal are zero. The best (uncoupled) designs can be represented by nxn diagonal matrices [2], [5], i.e., all entries off the main diagonal are zero. On the other hand, a coupled design is undesirable, because when a DP is modified, there is no effective solution for undesirable change on multiple FRs.

* Corresponding author.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 225–233, 2011.

Previous studies indicated the powerful function of AD may enhance product and process design abilities of the Research and Design department. Axiomatic design may help designers to structure and understand design problems [5]; however AD doesn't provide inventive principles or design suggestions to product designers [6], [7], [11], [12].

Theory of Inventive Problem Solving was developed by Genrich Altshuller and his colleagues, and is now being developed and practiced throughout the world. "TRIZ" is the acronym for this theory in Russian [1], [10]. TRIZ is a dialectic way of thinking in finding a suitable solution while facing a design predicament [4], [8]. The contradiction matrix of TRIZ provides designers which of the 40 inventive principles have been used most frequently to solve a problem that involves a particular contradiction [3], [9].

Therefore, unlike AD, TRIZ can provide concrete design suggestions. However, TRIZ does not locate design problems based on users needs like AD does. The purpose of this research intends to integrate the strengths of AD and TRIZ. This study wants to establish a systematic product design model by adopting some major tools from AD and TRIZ such as functional requirements, design parameters, design matrix, contradiction matrix and inventive principles. In addition, the proposed model's efficiency is analyzed and evaluated by a case study of a Handheld GPS product.

2 Methods

The proposed product design approach is shown in Figure 1 and briefly introduces in the following sections.

The first step is to analyze consumer needs of an existing or new product by AD's FRs and DPs. After completion of drawing the hierarchical diagrams of FRs and ADs, a design matrix may build based on the relationships between FRs and ADs. The items which need to be decoupled or uncoupled on a design matrix indicate the design issues to be addressed to increase usability.

Second, designers may consider these design issues and examine related engineering parameters one by one. The engineering parameters want to improve and may get worsened can locate design contractions by TRIZ's contradiction matrix. Then, designers may find out and adopt suggested inventive principles from TRIZ's contraction matrix to redesign or design products.

Finally, designers reconfirm the relationship between FRs and DPs of a new design on a design matrix until the contradiction is solved. In other words, a more decoupled or uncoupled design matrix represents this design solution with higher usability and users satisfactory.

The above procedures of the proposed model's efficiency is analyzed and evaluated by a case study of a hand-held satellite omniselector (Model number: GPSmap 60CSx) of the Garmin Corporation in Taiwan (http://www.garmin.com.tw/) (see Figure 2). Detailed description is illustrated in the Results section.

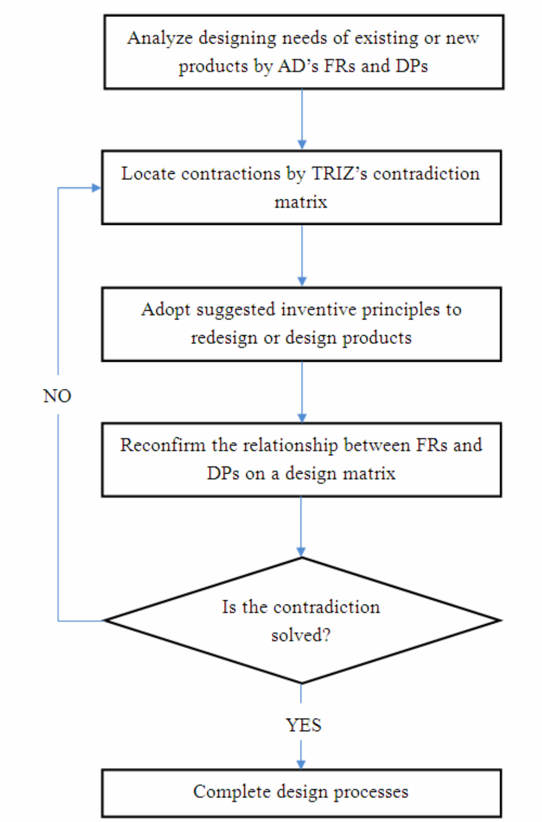


Fig. 1. The flowchart of our proposed product design model

Fig. 2. The satellite omniselector (GPSmap 60CSx)

3 Results and Discussions

In order to examine our proposed product design model, a case study of a hand-held satellite omniselector (GPSmap 60CSx) is conducted.

First, we analyze the main functions needed while using this product and we obtain nine main functions as below and the hierarchical diagram of the functional requirements is shown in Figure 3.

- FR1: Adjust map size, including zoom in (FR11) and zoom out (FR12).
- FR2: Power switch, including power on (FR21) and power off (FR22).
- FR3: Adjust display brightness, including brighter (FR31) and darker (FR32).
- FR4: Select functions, including function confirmation (FR41), quit (FR42) and cursor movement (FR43).
- FR5: Switch to main interface.
- FR6: Mark current position, including position confirmation (FR61) and Chinese characters input (FR62).
- FR7: Search destinations.
- FR8: Call sub-functions.
- FR9: Navigation.

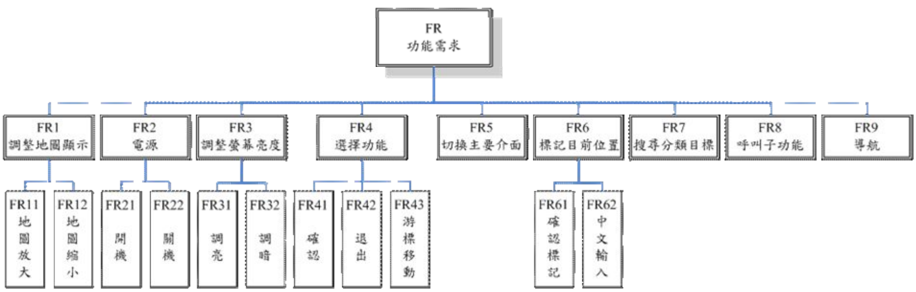


Fig. 3. The hierarchical diagram of the functional requirements (FRs) of the satellite omniselector (GPSmap 60CSx)

Next, this study analyzes the design parameters (DPs) of the current omniselector. There are ten control buttons on it as shown in Figure 4. Some control buttons may only use “press and release” to trigger their functions such as Zoom in (DP1), Zoom out (DP2), Find (DP3), Enter (DP6), Mark (DP7), Page (DP8), Quit (DP9) and Menu (DPa) keys. Both “press and release (DP41)” and “press and hold (DP42)” may use to initiate different functions of the Power key (DP4). “Press up (DP51)”, “press down (DP52)”, “press left (P53)” and “press right (DP54)” of the Rocker key (DP5) indicate the movement of a cursor to up, down, left and right directions. The hierarchical diagram of the design parameters is shown in Figure 5.

Afterwards, the functional requirements are mapping to the design parameters and the design matrix of original satellite omniselector is obtained as shown in Table 1. We can recognize from this design matrix that FR62 (Chinese characters input) and FR9 (interface in navigation) need to be decoupled to improve usability.

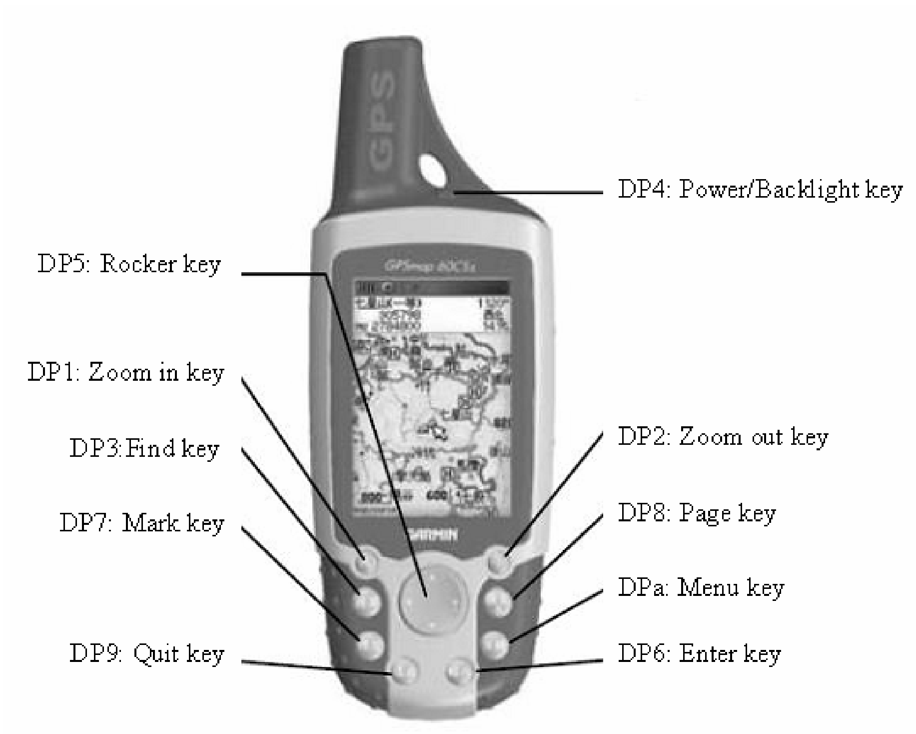


Fig. 4. The control buttons of the satellite omniselector (GPSmap 60CSx)

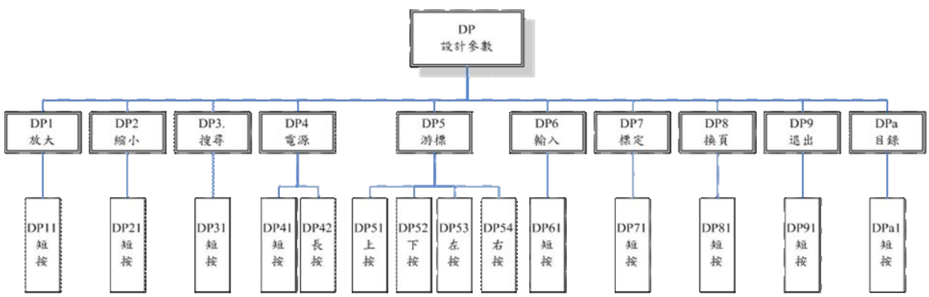


Fig. 5. The hierarchical diagram of the design parameters (DPs) of the satellite omniselector (GPSmap 60CSx)

Hereafter, we consider while redesigning FR62 and FR9, the features want to improve and the features may get worsened in terms of TRIZ's 39 engineering parameters. Then, the improved features and worsened features can be used to find out the suggested inventive principles on TRIZ's contraction matrix. Table 2 lists all the improved features, worsened features and inventive principles generated.

Table 1. The design matrix of the satellite omniselector (GPSmap 60CSx)

	DP11	DP21	DP42	DP41	DP61	DP91	DP51	DP52	DP53	DP54	DP81	DP31	DPa1	DP31
FR11	X													
FR12		X												
FR21			X											
FR22			X											
FR31				X										
FR32				X										
FR41					X									
FR42						X								
FR43							X	X	X	X				
FR5											X			
FR7												X		
FR8													X	
FR61														X
FR62					X	X	X	X	X	X				
FR9					X	X	X	X	X	X		X	X	

The suggested inventive principles are then used to stimulate design ideas. For example, principle #2 (extraction) implies we may separate navigation function from other functions to reduce user confusion.

Table 2. The suggested TRIZ's inventive principles while redesigning FR62 and FR9

Improved features	Worsened features	Inventive principles
6. Area of a stationary	12. Shape	-
	32. Manufacturability	40,16
	36. Complexity of a device	1,18,36
9. Speed	12. Shape	35,15,18,34
	29. Accuracy of manufacturing	10,28,32,25
	32. Manufacturability	35,13,8,1
26. Amount of substance	6. Area of a stationary	2,18,40,4
	12. Shape	35,14
	32. Manufacturability	29,1,35,27
	36. Complexity of a device	3,13,27,10
33. Convenience of use	29. Accuracy of manufacturing	1,32,35,23
35. Adaptability	20. Energy spent by a moving object	-
	22. Loss of energy	18,15,1
	29. Accuracy of manufacturing	-

After comparing feasibility, costs and convenience among design ideas, we decide to add two design parameters (DPb1 and DPc1) to decouple FR62 and FR9. The design matrix after redesign is shown in Table 3.

Table 3 is a decoupled design matrix. Even though a decoupled design is worse than an uncoupled one, it still allows the exact adjustment of the functional requirements. Especially Table 3 is obviously better than Table 1 in terms of usability. Our results indicate that the proposed product design approach can find out the usability problems as well as solutions efficiently.

Table 3. The design matrix of the satellite omniselector (GPSmap 60CSx) after redesign

	DP11	DP21	DP42	DP41	DP61	DP91	DP51	DP52	DP53	DP54	DP81	DP31	DPa1	DP31	DPb1	DPc1
FR11	×															
FR12		×														
FR21			×													
FR22			×													
FR31				×												
FR32				×												
FR41					×											
FR42						×										
FR43							×	×	×	×						
FR5											×					
FR7												×				
FR8													×			
FR61														×		
FR62					×	×									×	
FR9														×		×

4 Conclusions

This research demonstrates that we can locate design problems related to usability by AD's methods and generate design ideas from TRIZ's inventive principles. By combining these two theories, the proposed product design approach can find out the usability problems as well as solutions efficiently. When applying the proposed model to product design or new product development may avoid the cost waste and increase the design efficiency and usability during the product design and development processes.

Acknowledgments. This research is financially supported by National Science Council of Taiwan under contract number NSC99-2221-E-159-017.

References

1. Altshuller, G., Shulyak, L.: 40 principles: TRIZ Key to Innovative. Technical Innovation Center, Worcester (2002)
2. Axiomatic Design Solutions, Inc. (2011), `http://www.axiomaticdesign.com`
3. Conradie, I., Consultores, I., Chile, S.: TOC and TRIZ: Using a dual-methodological approach to solve a forest harvesting problem. TRIZ J. (2005)
4. Domb, E., Kowalick, J.: How to Bring TRIZ into Your Organization. TRIZ J. (1997)
5. Kulak, O., Kahraman, C.: Fuzzy multi-attribute selection among transportation companies using axiomatic design and analytic hierarchy process. Infor. Sci. 170, 191–210 (2005)
6. Lee, K.W., Ahn, Y.J.: Mutual Compensation of TRIZ and Axiomatic Design. TRIZ J. (2006)
7. Mann, D.: Axiomatic Design and TRIZ: Compatibilities and Contradictions. TRIZ J. (1999)
8. Mann, D.: Systematic Win-Win Problem Solving in a Business Environment. TRIZ J. (2002)
9. Rantanen, K., Domb, E.: Simplified TRIZ: new problem-solving applications for engineers & manufacturing professionals. CRC Press LLC, Boca Raton (2002)
10. Terninko, J., Zussman, A., Zlotin, B.: Systematic innovation: an introduction to TRIZ, pp. 65-89. St. Lucie Press (1998)
11. Yang, K., Zhang, H.: A Comparison of TRIZ and Axiomatic Design. TRIZ J. (2000)
12. Yang, K., Zhang, H.: Compatibility Analysis and Case Studies Axiomatic Design and TRIZ. TRIZ J. (2000)

User Characteristic-Based Information-Providing Service for Museum with Optical See-Through Head-Mounted Display: Does It Evoke Enthusiasm?

Yuki Yasuma and Miwa Nakanishi

Keio University, Fac. of Science & Technology, Dept. of Administration Engineering
Hiyoshi 3-14-1, Kohoku, Yokohama 223-8522, Japan
yasuma@a8.keio.jp, miwa_nakanishi@ae.keio.ac.jp

Abstract. In psychology, users' enthusiasm for products or services is categorized as a kind of intrinsic motivation. One theory states that enthusiasm is evoked when users perceive an adequate gap between their own characteristics and those of an object from the viewpoints of emotion, cognition, and ability. This study develops a method for computing an adequate psychological gap based on the characteristics of each user. We experimentally produce a service that makes each user feel the effect of the gap, and conduct a scientific evaluation. In particular, by focusing on the case of a museum, this study constructs an application to provide different sets of information to enable each user experience an adequate psychological gap with an optical see-through head-mounted display (OSD), and effectively evaluates whether this evokes user enthusiasm.

Keywords: Enthusiasm, Optical See-Through Head-Mounted Display, Museum, Information Providing Service.

1 Introduction

Different types of user interfaces are being presently developed and are expected to be applied to next-generation services. For these new types of services to be accepted in our society, practically utilized, and sustainably enjoyed, it is necessary that they make a positive impact on users' minds. That is, users should want to use the services, want to re-use them in other situations, tell others about them, and not forget them.

In psychology, users' enthusiasm for products or services is categorized as a kind of intrinsic motivation [1]-[10]. One theory states that enthusiasm is evoked when users perceive an adequate gap between their own characteristics and those of an object from the viewpoints of emotion, cognition, and ability [11] [12].

However, different users have different characteristics. Accordingly, to create products or services that evoke enthusiasm, it is necessary to provide an effect that corresponds to each user. Thus, this study develops a method for computing an adequate psychological gap based on the characteristics of each user. We experimentally produce a service that makes each user feel the effect of the gap, and conduct a scientific evaluation. In particular, by focusing on the case of a museum, this study constructs an

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 234–242, 2011.

application to provide different sets of information to enable each user experience an adequate psychological gap with an optical see-through head-mounted display (OSD), and effectively evaluates whether this evokes user enthusiasm.

2 Experiment

The gap between a user's own characteristics and those of an object from the viewpoints of emotion, cognition, and ability cannot be observed directly, and creating a method to measure it will require much examination. Thus, in this study, for the present step, we regard users' characteristics in the multiple aspects of emotion, cognition, and ability as simply a kind of inner characteristic. With this in mind, we designed the following experiment.

2.1 Experimental Environment

We built a mock museum booth with seven kinds of paintings. Fig. 1 shows the layout of the museum booth. The basic information on the seven kinds of paintings exhibited in the booth is summarized in Table 1. Each painting was framed and had a caption board next to it that provided basic information such as titles and artist names. Moreover, a preparatory room was prepared adjoining the museum booth where participants were given the informed consent form before the experiment and wrote down their reflections afterward.

2.2 Experimental Procedure

First, participants entered the preparatory room. Next, after agreeing to participate in the experiment, they were given fourteen items that are of interest to young people (selected based on [13]; see Table 2), and asked to classify them into five levels. The participants then selected the item that they were most interested in as the first level, the next two items as the second level, the next three items as the third level, the next four items as the fourth level, and the next four items as the fifth level, according to their own intuition. Here, the item in the first level was regarded as the one that had no gap from a participant's inner characteristics, while the items in the fourth and fifth levels were regarded as those with a large gap. Items in the second and the third levels were regarded as the ones with an adequate gap.

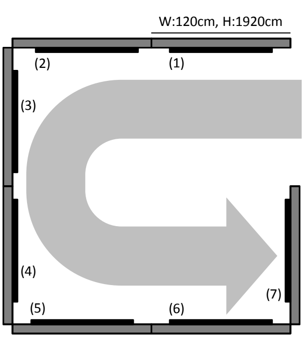


Fig. 1. Layout of the museum booth

Table 1. Paintings exhibited in the museum booth

No.	Title:	Artist:	Year:
(1)	Self portrait	Henri de Toulouse-Lauterec	1880
(2)	Die Gesandten	Hans Holbein der Jüngere	1533
(3)	Vulcan's Forge	Luca Giordano	1660
(4)	The Tower of Babel	Pieter Bruegel de Oude	1563
(5)	The School of Athens	Raffaello Santi	1511
(6)	Waterfall	Maurits Cornelis Escher	1961
(7)	Ultima Cena	Leonardo da Vinci	1498

Table. 2. Items of interest given to the participants

Sports	Travel
Music	Art
Game	Learning
Outdoor	Manga
Photo	Eating/Drinking
Reading	Internet
Movie	Gamble

After that, participants were asked to relax while sitting on the seat for three minutes. Next, they were explained about their upcoming experience in the museum booth. The main points of the instruction were the following: (a) After participants entered the museum booth, they could appreciate each painting by walking along the regular route (see Fig. 1); (b) After completing a circuit of the museum booth, participants were free to walk inside and appreciate any painting they liked; and (c) Participants could exit the museum booth any time they wanted to after appreciating the paintings. Finally, they were told to relax and enjoy the experience.

While the participants were given this explanation, a set of additional information for each one was edited as follows. Eight to twelve kinds of explanatory movies belonging to different genres were produced for each exhibit as elements of a set of information provided to each participant. These genres mostly corresponded to the interest items used for identifying participants' inner characteristics. One movie belonging to the genre corresponding to the interest items with an adequate gap from the participant's inner characteristics was chosen for each of the exhibits and summarized as a set of information provided to the participant. In case there were two or more movies belonging to the genre corresponding to the interest items with an adequate gap, one was chosen randomly. In this way, the movies were compiled into a program and installed on the mobile PC used for outputting the image to the OSD. An outline of one of the additional information sets prepared for "The Tower of Babel" (Pieter Bruegel de Oude, 1563) is shown in Fig. 2 as an example.

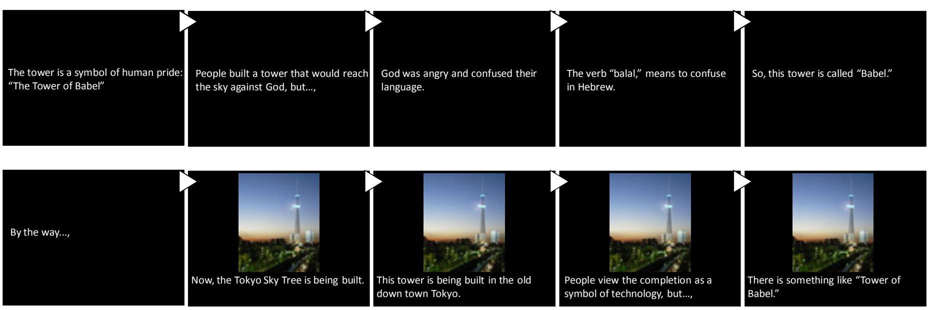


Fig. 2. Example of additional information (for "The Tower of Babel")

Participant put on the OSD before entering the museum booth. After checking visibility using a test pattern, they switched the image on the OSD to a standby image and entered the museum booth. After entering, the participants viewed each painting, referring to the additional information presented on the OSD. The additional information corresponding to one painting was a movie, which lasted around fifty to seventy seconds, some of which comprised multiple chapters. When

Fig. 3. A participant appreciating a painting

the participant wanted to advance to the next chapter, return to the beginning of the chapter, or skip the chapter, they could do so using a handheld mobile controller. During each participant's experience, no other person entered the museum booth. Fig. 3 shows a participant appreciating a painting.

The participants exited the museum booth and then returned to the preparatory room to write their reflections on a self-evaluation form.

2.3 Data

Before the museum booth experience, the face of each participant who relaxed for three minutes in the preparatory room was captured by a digital video camera (Cyber-shot DSC-TX7, made by SONY). Moreover, the results of the classification of the fourteen interest items were recorded as data.

During the experience, a participant's movement within the museum booth was captured by a digital video camera (Cyber-shot DSC-WX1, made by SONY) positioned at the center of the 266-cm-high ceiling. Moreover, the participant's face was captured by an ultra small video camera (AGENT CAM, made by Agent Camera) positioned above the frame of "The Tower of Babel."

After the experiment, participants filled out a self-evaluation form that asked for their reflection on the experience. The self-evaluation form was structured based on our previous study, in which we categorized the causal factors that induce enthusiasm by referring to the relevant theories [14]-[17] in psychology (see [18]), which included eight questions to be answered by free description (Table 3).

Table 3. Self-evaluation form

Self Evaluation Form
Please answer the following questions by free description.
(1)Please describe how you felt before entering the museum.
(2)Please describe how you felt when you appreciated the first painting.
(3)Please describe how you feel now (after the museum experience).
(4)Please describe as much as you can remember about what impressed you in the paintings you appreciated.
(5)Please describe what you thought of during the museum experience as much as you can remember (daily life, past memory, knowledge, etc.)
(6)How would you describe today's experience to your friends and family? Please describe it freely.
(7)Is there something more you want to know about the paintings you appreciated, please describe it freely.
(8)How much can you pay for the museum you visited today?
That's all. Thank you.

2.4 Experimental System

The retinal scanning display (prototype, made by Brother Industries, Ltd.) was adopted as the OSD used for providing information to the participant in the museum booth. This see-through type OSD provides the user with a full-color image with a size of around 800 × 600 pixels. Moreover, a laptop (VAIO VGN-T70B, made by SONY) was used to send the image to the OSD, and a mobile controller (Sa-Shi-41, made by KOKUYO S&T) was used to operate the image on the OSD.

2.5 Experimental Conditions

To examine whether providing information that had an adequate gap from the participants' inner characteristics enhanced participant enthusiasm in the museum booth experience, we set the following three cases as the experimental conditions.

(i) No additional information is provided to the participant: In a real museum, it is not rare to find exhibits with only basic information displayed on caption boards. As in this case, the participant experiences the museum booth without wearing the OSD.

(ii) Additional information that is chosen randomly (without considering the gap from the participant's inner characteristics) is provided: Regardless of the participant's inner characteristics, the set of information provided is composed by choosing movies one by one for each painting. The participant experiences the museum booth referring to this through the OSD.

(iii) Additional information that has an adequate gap from the participant's inner characteristics is provided: A set of information is composed in accordance with the processes described in Section 2.1. The participant experiences the museum booth referring to this information through the OSD.

2.6 Participants

The participants comprised 27 adults aged 20 to 36. Thirteen participants experienced the museum booth under the conditions of case (i), 7 under those of case (ii), and 7 under those of case (iii).

3 Results

Our analysis of the behavioral, physiological, and psychological data yielded the following results.

3.1 Time of Experience

We analyzed the video data captured from overhead to examine the duration of participants' stay in the museum booth. Figs. 4a, 4b, and 4c show typical behaviors that demonstrate how long a participant in cases (i), (ii), and (iii) stayed at each point in the museum booth.

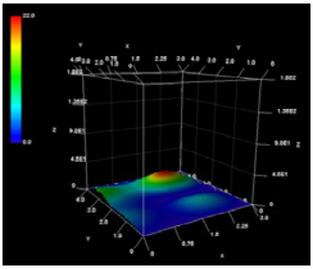


Fig. 4a. A participant's behavior in case (i)

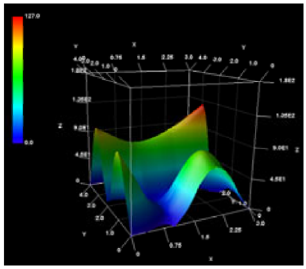


Fig. 4b. A participant's behavior in case (ii)

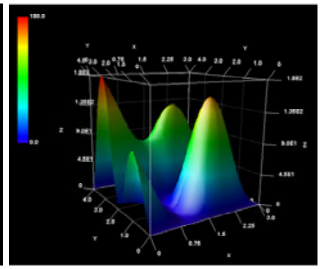


Fig. 4c. A participant's behavior in case (iii)

In case (i), when participants were given no information, the total time of stay in the museum booth was short. On the other hand, in cases (ii) and (iii), the participants spent significantly longer time in the museum booth. This is a natural result because the participants appreciated the paintings while referring to the additional information sets one by one. Further, by comparing cases (ii) and (iii), we find that the participants in case (iii) tended to stay longer in the museum booth than those in case (ii). This finding leads us to the following expectation regarding our primary question of whether providing additional information with a proper gap from participants' inner characteristics enhances their enthusiasm. That is, in case (iii), when consistent processes were applied to choose the additional information that had a proper gap from the participants' inner characteristics, the effect of enhancing enthusiasm would work on every participant while viewing any painting.

3.2 Eye-Blinking Frequency

Previous studies report that eye-blinking frequency changes when emotions related to enthusiasm, such as interest and those imparting satisfaction, are enhanced [19]. Therefore, we analyzed video data of the participants' relaxed faces captured before the museum booth experience and that of their faces as they appreciated "The Tower of Babel." Next, we calculated the difference in eye-blinking frequency between them. Fig. 5 shows the difference of the eye-blinking frequency of each case.

Eye-blinking frequency was significantly higher in case (ii) than in case (i) ($p < .05$) and in case (iii) than in case (ii) ($p = .11$). Because it is said that eye-blinking frequency particularly increases when the amount of interest and attention becomes more than steady, we can understand that the participant's enthusiasm was more enhanced in case (ii), when additional information was provided, than it was in case (i), when additional information was not provided, and further effectively enhanced in case

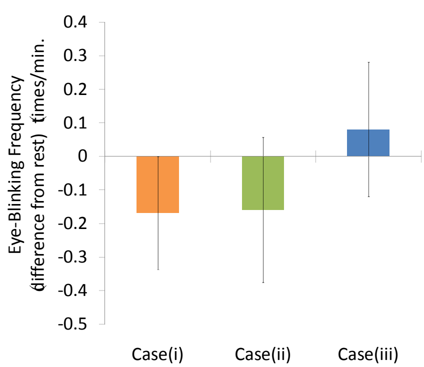


Fig. 5. Difference in eye-blinking frequency

(iii), when the additional information constructed in consideration of the participant's inner characteristics was provided compared to in case (ii), when additional information was constructed without this consideration.

3.3 Subjective Reflection

It is said that elaborate ideas and emotions emerge as a result of processes that induce enthusiasm in processes that induce enthusiasm [20] [21]. Thus, we analyzed the self-evaluation form completed by the participants after the museum booth experience and examined whether providing information involving a proper gap from the participant's inner characteristics helped participants in developing their ideas, thoughts, and emotions. In particular, we examined sentences that the participants used in their answers to eight questions (see Table 3). We divided them into the following three levels:

Level 0: Mentioning just the painting itself.
Level 1: Mentioning their knowledge or memory of the painting.
Level 2: Mentioning feelings or emotions based on their knowledge and memory of the painting.
Level 3: Mentioning a concrete desire to do something based on their feelings or emotions.

If a participant wrote two or more sentences, we classified each separately. Fig. 6 shows the result of classification of a total of 153 sentences. This chart also reflects the case where a sentence included multiple parts to be classified into different levels.

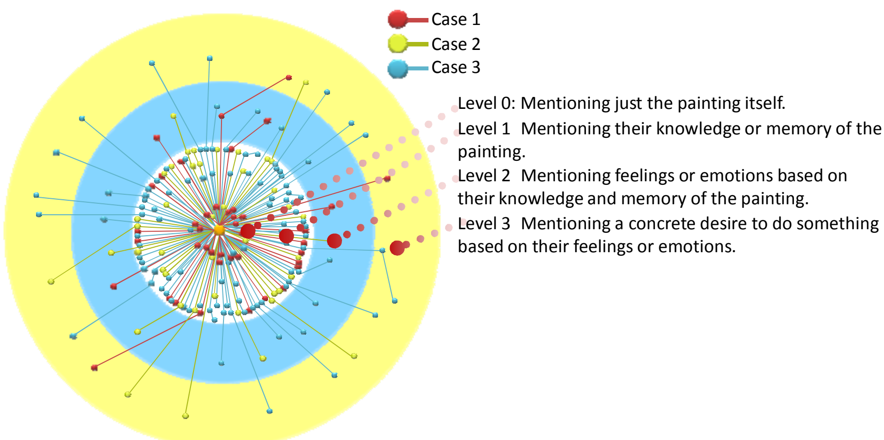


Fig. 6. Classification of the participants' self-reflection sentences into four levels

Focusing on the result of the classification of each case, we find the following characteristics. Most of the sentences written by participants after the museum booth experience under the conditions of case (i) were about primary impressions given by the paintings, and more than 40% were classified as level 0. On the other hand, for case (ii), more than 50% of the sentences mentioned the participant's own knowledge, memory, and experience associated with a painting, which corresponds to level 2. Furthermore, for case (iii), there were more sentences that included the desire for action inspired by interest in a painting than for the other cases. These results suggest that the participants' ideas and emotions generated through the museum booth experience became more spontaneous in case (ii) than in case (i) and in case (iii) than in case (ii). Based on the above, it was suggested that providing additional information on each painting made the museum booth experience more impressive, and that adding the effect of a proper gap to the additional information triggered development of users' enthusiasm.

4 Conclusion

This study experimentally examined a service for providing annotative information with an adequate gap based on each user's characteristics during a museum booth experience; moreover, it evaluated whether this experience evoked their enthusiasm from multiple viewpoints. The results of the experiment suggested that the service enabled users to enjoy the museum space more profoundly. Moreover, the results suggested that users' enthusiasm can be raised to a greater extent by applying a scientific method to compute an adequate gap in accordance with each person's characteristics.

In the present day scenario, when many well-developed products and services are available with no apparent differences among them, objects that can move people's hearts are of significant interest to those who provide products and services. Moreover, designing products or services in consideration of people's feelings and emotions in addition to functions and performance is a way to satisfy today's users.

Acknowledgement

We would like to thank Mr. Tomohiro Sato and Mr. Makoto Nomura (Brother Industries, Ltd) who provided helpful comments and suggestions.

References

1. Deci, E.L., Ryan, R.M.: Self-Determination Theory and the Facilitation of Intrinsic Motivation. Social Development, and Well-Being, American Psychologist 55(1), 68–78 (2000)
2. Deci, E.L.: Intrinsic motivation. Plenum Press, New York (1975)
3. Harlow, H.F.: Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. J. Comp. Physiol. Psychol. 43, 493–508 (1950)
4. Horlow, H.F.: Mice, monkeys, men, and motives. Psychological Review 60, 23–32 (1953)
5. Deci, E.L., Ryan, R.M.: Intrinsic Motivation and Self Determination in Human Behavior. Plenum Press, New York (1985)
6. Harter, S.: Effectance motivation reconsidered: Toward a developmental model. Human Development 21, 34–64 (1987)

7. Harter, S.: A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. Developmental Psychology 17, 300–312 (1981)
8. Berlyne, D.E.: Conflict and choice time. Brit. J. Psychol. 48, 106–118 (1957)
9. Vallerand, R.J., Bissonnette, R.: Intrinsic, extrinsic and motivational styles as predictors of behavior: A prospective study. J. Pers. 60, 599–620 (1992)
10. White, R.W.: Motivation reconsidered: The concept of competence. Psychological Review 66, 297–333 (1950)
11. Hunt, J. McV: Motivation inherent in information processing and action. In: Harvey, O.J. (ed.) Motivation and Social Interaction: Cognitive Determinants, pp. 35–94. Ronald, New York (1963)
12. Hunt, J. McV: Intrinsic motivation and its role in psychological development. In: Levine, D. (ed.) Nebraska Symposium on Motivation, vol. 13, pp. 189–282. University of Nebraska Press, Lincoln (1965)
13. Ministry of Internal Affairs and Communication: Survey of Time Use and Leisure Activities.: Table 60-1. Participants in Hobbies and Amusements by Sex, Usual Economic Activity and Age (2006)
14. Deci, E.L.: Intrinsic motivation, extrinsic reinforcement and inequity. J. Pers. Soc. Psychol. 22, 113–120 (1972)
15. Berlyne, D.E.: Conflict and choice time. Brit. J. Psychol. 48, 106–118 (1957)
16. Dember, W.N., Earl, R.W.: Analysis of exploratory, manipulatory, and curiosity behaviors. Psychological Review 64, 91–96 (1957)
17. Berlyne, D.E.: Structure and direction in thinking. John Wiley & Sons, Chichester (1965)
18. Yasuma, Y., Nakanishi, M., Okada, Y.: Can "tactile kiosk" attract potential users in public. In: Proceedings of the 3rd International Conference on AHFE (Applied Human Factors and Ergonomics), CD-ROM, Miami, USA (2010)
19. Hall, R.J., Cusack, B.L.: The measurement of eye behavior: Critical and selected reviews of voluntary eye movement and blinking. U.S. Army Technical Memorandum, pp. 18–72 (1972)
20. Horie, T., Inuzuka, A., Ikawa, Y.: The Effects of Intrinsic Motivation on Knowledge Creating Behavior. Journal of the Japan Techno-Economics Society 504, 66–71 (2009)
21. Miwa, K., Ishii, N.: Cognitive Approaches to Creative Activities. Journal of the Japanese Society for Artificial Intelligence 19(2), 1–9 (2004)

Human-Information Interactions with Complex Software

Michael J. Albers

East Carolina and University
Greenville NC 27858
albersm@ecu.edu

Abstract. This article extends the analysis of a usability test of C2PC, a US Marine Corps command and control software product. The study revealed the C2 operators were able to perform simple tasks, but had difficulty combining those simple tasks into realistic tasks. These differences highlight the need to both consider the complex interactions during HCI design and for using complex scenarios when testing complex systems. Poor human-information interaction (HII) is reflected in designs which fail to support effectively rolling up the individual tasks into the complex interactions that people must perform. Usability tests show basic tasks can be accomplished but these systems fail to support people solving open-ended, unstructured, complex problems which require extensive and recursive decision-making or problem solving This paper discusses how the issue appears in many software products, causing problems in effectively communicating information. It considers the broader design issues for complex information spaces.

Keywords: Human-information interaction, usability testing, complex information systems, complex system design.

1 Introduction

Individuals seek information because they realize that their knowledge about a situation is incomplete and that the information they need can be found within the system. Design teams are increasingly being called upon to address information needs which go beyond providing simple answers or step-by-step instructions and which involve communicating information for open-ended questions and problems. Questions and problems that can only be addressed by providing information specific to a situation and presenting it in a way that supports various users' goals, information needs, and cognitive processing strategies.

With respect to HCI and usability, it's the human-information interactions (HII) which lead to comprehending the information [8]. As a system's complexity or its information content complexity increases, the need to understand and predict people information interactions make it imperative that designs work from a basis of how people organize their thoughts. This, of course, contains implicitly the importance of understanding people's thoughts and perceptions.

Mirel's [15] studies have found users have different conceptions of how to accomplish a task, which she described by saying: "In actual work settings, users

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 245–254, 2011.

define their own tasks and task needs according to situational demands, not program design" (p. 15). The design of those systems must encompass a totality that revolves around the goals and information needs of a person and supplies information that makes sense within the person's real-world situation.

Handling the shift from a linear information model to the ill-structured model requires a design shift. Redish [19] argues that documentation move up a level and address goals and task repertoires and she recognizes that that emphasis on "higher than discrete task" is crucial for anyone doing everyday work. Accomplishing this requires us to view the creation and presentation of content and its subsequent communication from a humanistic viewpoint, rather than a mechanistic one [25]. The user goals and information needs must be placed within a proper social and technical contexts and designed to assist people, rather than doing it for them [21, 22]. The driving force for this shift arises because people, rather than machines, are reading the information. [3, p. 4]

The issues of simple and complex information systems and designs which either conflate the two or assume only simple interactions form the basis of many poorly designed systems [5]. I've considered the issue of simple and complex in more detail in earlier work [3]; here I'm using the same definitions, which in a reduced form are:

Simple systems. A single or multiple fully-defined paths from problem to solution can be defined. It is possible to state that an action was either correct or incorrect.

Complex systems. Multiple and ill-defined paths to a solution exist. The solution depends on situation-specific context and can change between people or across time. It is impossible to clearly state if an action is correct.

People within a complex situation need to react within a highly dynamic environment and are faced with open-ended, complex problems. In a complex situation, the problem will almost away include factors or circumstances not foreseen as part of the original analysis. In these situations, the analysis and design must consider the highly dynamic situational context of information, the aspects of the information, and the information interrelationships required to support fundamental user wants and needs. These fundamental issues apply to any dynamic complex environment, such as military command and control, healthcare or financial systems. How can the design issues be resolved and what does this say for future system design? What does it mean for HCI to transition from designing for simple operations to complex information-based interactions. How do well do common HCI design strategies scale to complex dynamic environments, how can we help support the scaling, and how do we test the quality of the resulting design. This paper explores these ideas.

1.1 Communication, Not HCI Performance

Hollnagel [11] has described human work as fitting on a scale ranging from "doing" to "thinking." A simple task design based on performing any single menu option (with menu used loosely to include simple web interactions) fall within the "doing" region of the scale: a person has a stable, linear path to complete a task. With the rise of software and web pages providing interactive support for problem-solving or providing large amounts of information which a person is expected to mentally process and act upon, the interaction has shifted into the "thinking" region. Here people are engaged in knowledge intensive “thinking” activities focused around diagnosis, planning, and problem solving, rather than manually “doing” tasks. This shift has fundamental consequences for the design, content, and usability testing of the resulting system and the information it must communicate to the user.

Many traditional usability tests focus on function (essentially the button pushing) and not on the process of how the information is used. Recent work in usability calls this model into question, arguing that it is too divorced from reality to provide useful information [18]. In its place a new approach has been called for [20], one that recognizes that most users operate, or carry out their tasks, within complex systems that present multidimensional challenges—layers of changing depth that, unfortunately, traditional usability methods often cannot adequately measure. [5, p. 3]

The system’s overall goal involves communicating information to a user. The development and testing of any design and interaction should focus on the quality of that communication. Do the users comprehend the information and can they apply it to their situation? Thus, a goal of a usability test in a complex situation is to ensure that the information is being efficiently communicated and that the person is building information relationships. Developing an understanding of a complex situation requires understanding the relationships between information elements, not simply interacting with single information elements. [6]. This is a departure from traditional usability testing methods which dealt with single tasks and measurements of easy to quantify data, such as time to complete a task or total mouse clicks.

1.2 Design and Usability of Complex Information Systems

A problem facing design teams is that as products get larger and more complex, usability testing methods prove increasingly inadequate for testing usefulness. This has recently received discussion in the literature [1, 3, 16, 18, 19], and most of us have experienced it when performing tests of larger systems. Traditional, one- or two-hour usability tests find nothing of significance; yet when the product is used in a real-world situation people fail to complete tasks successfully [9].

Part of the testing problem can be a desire (demand?), either by management or the test team, to produce numerical results. Focusing on individual components makes it relatively simple to construct quantitative measures such as time to complete a component task or total number of clicks. However, quantitative evaluation of a complex design is suspect since the quantitative evaluation often privileges the easy-to-measure over other factors without a solid theoretical foundation for that privileging [24]. Instead, design teams creating products for complex systems must embrace flexibility and usability, and work to ensure that flexibility meets user needs [23]. On the other hand, as I’ll discuss later, too much flexibility can also impact usability if the users don’t need or want that flexibility to achieve their goals.

This articles extends the analysis of the results of a usability test of C2PC, a US Marine Corps command and control software product, which has been published elsewhere [7]. First, I’ll briefly review the C2PC study. Then, I’ll consider three design factors arising from the study and consider how they apply across many different design scenarios.

2 C2PC Test Summary

C2PC, currently deployed by the US Marine Corps, is a Windows-based military command and control (C2) system which provides an integrated views of the battle space in the Command Operations Center (COC). The C2 operator must manipulate a complex set of information in order to maintain situation awareness of the battle space (figure 1). Commanders at multiple command levels, from the battalion up to the main force level, depend on C2PC to maintain situational awareness and make operational decisions. To support these commanders, the C2 operator must manipulate a complex set of information (tracking military units, movement paths, and targets) to maintain a tactical picture of the battle space. By design, C2PC requires the C2 operators' interaction to shift between the map and a set of changing injectors (essentially modes) to accomplish their goals.

A usability study [7] revealed an underlying design problem seems to be that the C2PC designers provided an interface for performing a sequence of individual simple tasks, while real-world operation requires working in a cyclic manner with a complex set of information focused around open-ended questions. Instead, both the usability test and reports by training center instructors, reveal C2 operators tend to work either within the injector or within the map; they do not fluidly shift between them. Only after their preferred interaction area fails to yield a solution do the C2 operators shift to the other area. In the usability test, multiple C2 operators failed to ever change the zoom level (4 of 13) or change the injector (3 of 13) from its original settings. As would be expected, these C2 operators had high failure rates on the usability test tasks.

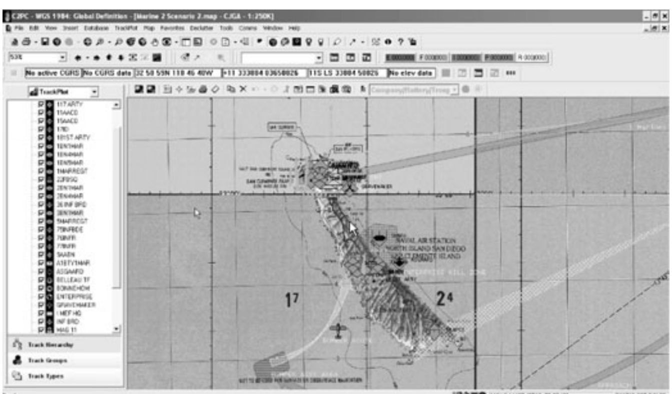


Fig. 1. C2PC main screen. A zoomable map of the battle space is on the right which shows all of the units. The list on the left is the injector, which varies depending on the software mode.

During the usability test the observations revealed an interesting disconnect between the underlying design assumptions and how the C2 operator interacted with the system. Many users never used the map zoom functions and had difficulty with simultaneously working with injectors and the map. A substantial usability problem which contributed to many task failures was the C2 operator's highly non-optimal use of the zoom function. Some subjects never changed the zoom and others made non-productive changes, such as zooming in too much or not enough, and then attempted to interact with C2PC. A second major usability problem was that C2 operators would search through the contents of the wrong injector before changing it or before going to the map to try and complete the task.

The problem with C2PC was neither the technical nor the computer-interaction level design, but rather that its HII aspects do not seem to have anticipated how people would actually interact with the program. Or more importantly, how they would interact with integrating all of the information available in the system and in their environment. A software design team too typically sees the software project as a collection of basic tasks, but realistic use requires combinations of those tasks (which the design team may or may not understand). The basic usability issues found seem to reveal that its design has not anticipated how people would actually interact with the program. Usability tests of individual tasks show they can be accomplished with minimal problems. However, individual interface tasks are not performed in isolation as people work with real-world tasks. Instead, they achieve their real-world goals by performing a grouping of interface-level tasks. C2PC fails to support effectively rolling up the individual tasks into the complex interactions which the C2 operator must perform. It fails to support working with solving open-ended, unstructured, complex problems which typically require extensive and recursive decision-making or problem solving [18, 19]. In the bigger picture, this support of individual tasks and lack of support for combing them into real-world tasks is a consistent problem with the design of many complex information systems[1].

Granted, the ideas of the previous paragraph are well accepted in HCI research and design; however, the implementation issues surrounding them have profound effects on people's actual performance and remain an open research question.

3 Broader Implications of the Results C2PC Usability Test

The remainder of this article will not dwell on the results of the C2PC usability test, but will consider how the fundamental HII disconnect found within that study occurs across many software products and causes an excess of failures in effective information communication. In an era long past, writing technical information meant writing for closed-ended questions, such as procedures or answering questions of "what is X." (Whether or not that was effective communication is a different question.) Thinking in terms of simple tasks or answering closed-ended questions exposes substantial issues with the usability of many designs. As Howard [13] summed it up: "Complex problems for users can masquerade as simple problems for authors, editors, designers, and even expert reviewers with experience in the tasks being performed" (p. 84).

Consider how, rather than answering "what is X," many texts and web pages are created with an intent of providing complex information for open-ended questions with a goal of communicating information that address questions such as "How are X and Y effecting Z?" Modern design teams need a much deeper understanding of what constitutes effective communication. Design teams are increasingly being called upon to address information needs which go beyond providing simple answers or step-by-step instructions and which involve communicating information for open-ended questions and problems. Questions and problems that can only be addressed by providing information specific to a situation and presenting it in a way that supports various users' goals, information needs, and cognitive processing strategies.

The problems found in the C2PC study tended to be complex problems which the design teams had viewed a collection of simple problems. However, complex problems are not the sum of their parts, they are much more than the sum. This section examines three design issues which expand on this idea.

3.1 Different Audiences Exhibit Radically Different Interaction Strategies

The C2PC study revealed interesting disconnects between the underlying design assumptions and how C2 operators at lower experience levels interacted with the system. They were able to perform simple tasks ("doing"), but were unable to combine those simple tasks into realistic tasks ("thinking"). These differences highlight the need to both consider the complex interactions during HCI design and for using complex scenarios when testing complex systems [19].

With complex information, the information needs and effective presentation formats can radically vary between audience groups. The system is expected to effectively communicate high-quality information to each group, although the information needs and background knowledge vary with each group.

In C2PC, the users could work with the system without using the zoom function. The C2 operators knew how to change the zoom, but they made a conscious decision not to do it or to minimize the changes. As a result, they were very inefficient and error-prone, but they could accomplish the task (at least they believed they had accomplished the task). With experience, they started to use zoom. From a designer's viewpoint, it's hard to visualize any screen which is not already at the proper zoom level or on the proper injector, or which the user cannot easily adjust with at most a couple of mouse clicks. And this is a major issue: a designer visualizes how to perform the task properly and has trouble figuring out how to cue the user that something needs to change (especially without interfering with expert level performance). The beginner or low-intermediate user, on the other hand, does not perform the expected integrated interaction and has usability problems which they often associate with bad design.

Most systems have to deal with a range of experience levels. As people's experience with a system increases, their interaction strategies change. The too common mantra "Write for the novice and the expert will understand it too" is not true; providing information designed for the lowest knowledge audience group impairs comprehension in the other groups [14]. The changes in interaction strategies are not simply becoming better or faster, but changes in how the overall system interaction is viewed and how individual tasks are mentally assembled.

Satisficing plays a major role with how people interact with a system. "Users will make sense of a document in whatever ways best suit their needs and purposes at the time; they will usually settle for the first solution they find that satisfactorily answers their particular problem at that point in time" [12, p. 69]. This also means their performance changes can occur in jumps as they eventually are forced to replace a method with worked with a new method which handles a broader set of circumstances. As example of satisficing, consider a situation which many of us would claim is simple, but which was found to be otherwise: looking up citation formatting in a writing handbook. Students stopped at the first piece of information which looked good and never evaluated if the format they had picked fit their needs [13]. From a design aspect, the text seemed very straightforward. From a complex usability aspect, the information was not presented in a manner which supported the student in a problem-solving situation.

These same issues are seen in using electronic medical record systems. There is a huge amount of available information and each person who interacts with it needs a different subset of that information. As a result, how they interact with it differs by audience. "A human/computer interface has to be ergonomic (coherent, concise, reactive, structured and flexible) and customizable to and by the user. It must automatically adjust its look and feel to suit the requirement of individuals or groups of users" [10, p. 220].

3.2 Complex Situations Revolve around Decision Making

The operator-based HCI psychology research has defined a good set of guidelines on how to effectively communication simple information. However, those same guidelines can work at cross-odds with communicating complex information. The question is how accepted design practices scale for a complex application such as C2PC. Highly effective interactions with these complex applications are more than simply a sum of a sequence of individual tasks. Instead, the interaction requires a constant back and forth between multiple interface elements with the current best answer constantly changing in the dynamic information environment. Examining a system for genome data analysis, Dicks [9] said:

The product was better designed than most such systems, and it included many of the necessities for successfully completing such work that Mirel [18] points to, including means for nonlinear, multipath studies; for collaboration among disparate professional groups….Nonetheless, while each of the smaller parts seemed to work quite well, it proved extremely difficult to determine how useful the system was in an overall sense (p. 209).

In these situations, the information is essentially dumped before the user, who was left to figure out what to do with it. Unfortunately, people are very poor at figuring out what to do with it since they inherently strive to minimize cognitive load and effort. With a design based around efficient performance of simple tasks, the HII imposes a high cognitive load since the person is dealing with both complexities of their goals, and the information in the situation, and the mental transformation required for interacting with the display.

Ensuring that information is usable rather than a dump, means that usability testing in complex situations needs to focus on knowledge-based measures which measure the internal thought processes. Typically this involves asking the person to describe the current situation at various points and asking what information led to those conclusions. Performance-based measures (task time, click counts, etc.) are easy to collect but provide little useful information [5].

3.3 HII Design Issues and Interaction Complexity

As Mirel says [17, p. 233], "Complex tasks and problem solving are different in kind not just degree from well-structured tasks." Complex information is multidimensional. There are no simple answers and there are a dynamic set of relationships which change with time and in response to situational changes [3].

Ensuring proper information salience should be a significant focus of design teams when dealing with complex information. It is easy to create designs were all of the information receives the same level of emphasis, but this results in designs which are difficult to interact with. The user now has to read an excess of information (since the irrelevant or lower importance information is not flagged) and mentally reduce the information. Experts in an area can accomplish this, it's one of the distinguishing characteristics of an expert, but both beginner and intermediate uses cannot. As a result, the person's contextual awareness suffers [4] and they are unable to build the necessary information relationships to fully grasp the situation [6].

Interestingly, the person may understand what is happening in the system. In the C2PC study, the Marines were able to give a basic explanation of how their task operated within C2PC. However, when given specific tasks to accomplish, they were unable to effectively interact with the interface to perform that task.

4 Conclusion

Traditionally, HCI addressed highly structured situations, with a basic goal of efficiently completing a task. But many interaction situations are no longer highly structured; instead, they have shifted to complex situations revolving around information seeking, problem-solving, and decision making, which call for developing a high level of contextual awareness. This type of work needs a deep understanding of HII, which is essential to designing interfaces that communicate complex information.

Poor HII is reflected in designs which fail to support effectively rolling up the individual tasks into the complex interactions that people must perform. Usability tests show basic tasks can be accomplished with minimal problems. But any complex system must provide support for people working with solving open-ended, unstructured, complex problems which typically require extensive and recursive decision-making or problem solving [2]. Any complex information system involves complex situations; design teams must consider what factors influence how people perceive the information in their contextual environment and then build on those perceptions to enable the selection of relevant information to support judgments, decisions, and actions.

This paper looked at a study of C2PC and extended the ideas to other systems. A fundamental interaction problem with the design of complex systems is that people often fail to both effectively and efficiently work with the system. Across these complex software platforms, performance of basic operations tend to be very quick and simple to learn. But the learning curve for working within a complex environment is much more than simply knowing what buttons to push. An open research question revolves around designing easy to learn complex systems that remain easy to use and that support a multitude of interaction styles. The interaction moves from "doing" the button pushing to "thinking" about which buttons to push, in which order, why to push them, and how to support evaluation of the result of the push. Software does not exist to provide people with button pushing experiences; it exists to help people understand a situation and make decisions which influence that situation.

Acknowledgements. The C2PC research was performed while the author was on an Office of Naval Research Summer Facility Fellowship at Naval Undersea Warfare Command, Newport, RI.

References

1. Albers, M., Still, B. (eds.): Usability of Complex Information Systems: Evaluation of User Interaction. CRC Press, Boca Raton (2010)
2. Albers, M.: Complex problem solving and content analysis. In: Albers, M., Mazur, B. (eds.) Content and Complexity: Information Design in Software Development and Documentation, pp. 263–284. Erlbaum, Mahwah (2003)
3. Albers, M.: Communication of complex information: User goals and information needs for dynamic web information. Erlbaum, Mahwah (2004)
4. Albers, M.: Design for Effective support of user intentions in information-rich interactions. Journal of Technical Writing and Communication 39(2), 177–194 (2009)
5. Albers, M.: Usability of complex information systems. In: Albers, M., Still, B. (eds.) Usability of Complex Information Systems: Evaluation of User Interaction. CRC Press, Boca Raton (2010)
6. Albers, M.: Usability and information relationships: Considering content relationships when testing complex information. In: Albers, M., Still, B. (eds.) Usability of Complex Information Systems: Evaluation of User Interaction. CRC Press, Boca Raton (2010)
7. Albers, M.: Design and usability: Beginner interactions with complex software. Journal of Technical Writing and Communication (in press)
8. Albers, M.: Human-Information Interaction and Technical Communication: Concepts and Frameworks. IGI Global (in press)
9. Dicks, S.: Designing usable and useful solutions for complex systems: A case study for genomics research. In: Albers, M., Still, B. (eds.) Usability of Complex Information Systems: Evaluation of User Interaction. CRC Press, Boca Raton (2010)
10. Ghedira, C., Maret, P., Fayn, J., Rubel, P.: Adaptive user interface customization through browsing knowledge capitalization. I. J. Medical Informatics 68(1-3), 219–228 (2002)
11. Hollnagel, E.: CREAM—Cognitive reliability and error analysis method (2006), http://www.ida.liu.se/~eriho/CREAM_M.htm
12. Howard, T., Greer, M.: Innovation and Collaboration in Product Development: Creating a New Role for Usability Studies in Educational Publishing. In: Albers, M., Still, B. (eds.) Usability of Complex Information Systems: Evaluation of User Interaction. CRC Press, Boca Raton (2010)
13. Howard, T.: Unexpected Complexity in a Traditional Usability Study. Journal of Usability Studies 2(3), 189–205 (2008)
14. McNamara, D.S.: Reading both high and low coherence texts: Effects of text sequence and prior knowledge. Canadian Journal of Experimental Psychology 55, 51–62 (2001)
15. Mirel, B.: Analyzing audiences for software manuals: A survey of instructional needs for real world tasks. Technical Communication Quarterly 1(1), 15–35 (1992)

16. Mirel, B.: Advancing a vision of usability. In: Mirel, B., Spilka, R. (eds.) Reshaping Technical Communication, pp. 165–187. Lawrence Erlbaum, Mahwah (2002)
17. Mirel, B.: Dynamic usability: Designing usefulness into systems for complex tasks. In: Albers, M., Mazur, B. (eds.) Content and Complexity: Information Design in Software Development and Documentation, pp. 233–261. Erlbaum, Mahwah (2003)
18. Mirel, B.: Interaction Design for Complex Problem Solving: Developing Useful and Usable Software. Morgan Kaufmann, San Francisco (2003)
19. Redish, J.: nderstanding readers. In: Barnum, C., Carliner, S. (eds.) Techniques For Technical Communicators, pp. 15–41. Macmillan, New York (1994)
20. Redish, J.: Expanding usability testing to evaluate complex systems. Journal of Usability Studies 2(3), 102–111 (2007)
21. Ruskin, J.: The Humane Interface: New Directions for Designing Interactive Systems. Addison-Wesley, New York (2000)
22. Schriver, K.: Dynamics in Document Design: Creating Texts for Readers. Wiley, New York (1997)
23. Shearer, H.: An Activity-theoretical approach to the usability testing of information products meant to support complex use. In: Albers, M., Still, B. (eds.) Usability of Complex Information Systems: Evaluation of User Interaction. CRC Press, Boca Raton (2010)
24. Sutcliffe, A., Karat, J., Bodker, S., Gaver, B.: Can we measure quality in design and do we need to? In: Proceedings of DIS 2006: Designing Interactive Systems: Processes, Practices, Methods, & Techniques, pp. 119–121 (2006)
25. Whitburn, M.: The ideal orator and literary critic as technical communicators: An emerging revolution in english departments. In: Lundsford, A., Ede, L. (eds.) Essays in Classical Rhetoric and Modern Discourse, pp. 230–248. Southern Illinois University Press, Carbondale (1984)

The Importance of Rigor in Usability Studies

Randolph G. Bias

School of Information
The University of Texas at Austin
Austin, TX 78701
rbias@ischool.utexas.edu

Abstract. Well-designed and conducted usability efforts – "big U Usability including the entire user-centered design process, from requirements gathering through prototype creation and evaluation, through system usability testing, through delivery/cut-live, and to field testing and beyond, and even fundamental usability research – are vital to the creation of human-information systems that actual people can use to carry out their intended tasks. Done poorly, such efforts can be worse than nothing. We have crafted this conference session, "Usability Studies: Rigor or Rigor Mortis?," to highlight good examples of rigor applied to usability studies, and to illustrate how important a rigorous approach is.

Keywords: Usability, user-centered design, empiricism, rigor.

1 Introduction

Our field has gone by many names, names denoting and connoting largely overlapping disciplines: human factors, usability engineering, user-centered design, interaction design, user experience design. The main tenet that we all share – the one thing that is certain to be at the center of the Venn diagram that would represent all of our cognate fields – is a desire to design artifacts to fit the user. The field of usability is replete with methods allowing us to collect user data, data to inform (early on in the development cycle) and later validate our designs. Plus, the field of usability is undergirded by research in human information processing, fundamental research on how human beings take in and process information.

Usability engineering founding father Jakob Nielsen (2005), says "we will need to scale up by a factor of 100 in terms of available personnel" (p. 3). He goes on to say that we need to ". . . expand usability beyond the usability professionals. If everybody needs usability, then everybody should do usability. We may need 50 million people who know some usability, but [that] is easier to achieve than getting one million people who are full-time usability experts" (p. 3).

Usability professional extraordinaire Steve Krug admonished us to *Don't make me think* (2005), and then went on to offer *Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems* (2010).

If we are in such desperate need of usability support, and if it so easy that someone with no other training might read one thin (though outrageously well written and

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 255–258, 2011.

illustrated) book, or perhaps take a short course, and then be (or at least feel) empowered and armed to conduct usability studies, isn't this a good thing? Isn't it the case, as many software developers and some usability professionals have argued, that some usability testing is better than none, no matter how it is carried out? Certainly, on the face of it, usability is neither brain surgery nor rocket science. But in this paper I would like to argue, and other papers in this session will support, that some usability support is worse than nothing.

2 Three Arguments

2.1 It's Not Rocket Surgery, But It's Not Back-Yard Deck Building, Either

Elsewhere I've written on what I perceive to be the dangers of amateur usability engineering (Bias, 2003). Briefly, if you do a poor job of writing your computer program, this gets "discovered" in system test, or quality assurance testing (usually). If you do a poor job of usability engineering, this fact may not be discovered until the product has shipped or "gone live," and either the customer support phones start ringing off the hook or visitors start leaving your web site in droves for no apparent reason.

Usability engineering may be a more accessible discipline than medicine or rocketry, but it is more complicated than most do-it-yourself projects. It is replete with history, best practices, and methods that can be done well or poorly, and that can be applied at good or poor times in the design and development cycle. To paraphrase Abraham Maslow, "To the usability practitioner who only has a usability lab, every usability situation he/she encounters begins to look to call for an end-user test." There are many user interface evaluation methods that might be employed, and lab testing is just one of them. Selecting the best method requires training and experience; selecting a suboptimal method may lead to disaster, and can certainly lead to reduced return-on-investment (ROI) for a team's usability hour and dollar invested.

The same is true with human information processing research. Inferences made from poorly designed studies run the risk of (poorly) informing practice for years to come.

In advance of our debate on the dangers of amateur usability engineering at the 2008 Usability Professionals' Association annual meeting (see Bias, 2008), we handed out buttons for people to pin on their jackets, buttons arguing the pro and con case for "some, any" usability engineering support. My favorite was "Do I have to stay for the whole day of the tutorial to be a usability professional?" To do a good job of usability research or practice requires more than good intentions and a sensitization to the value of usability.

2.2 The Problem of Credibility

When that aforementioned, imaginary poor job of usability engineering support gets done, the vice president who paid for it does not think "We received poor usability support." He/she thinks "Usability isn't worth it." Ours is still a young discipline, scrambling for a routine seat at the web-and-other-software-design table (e.g., Bias &

Mayhew, 1994). There is no widely-accepted certification of usability professionals. For our discipline to allow (empower, encourage, endorse) less-than-professional work only retards our progress as being seen as professional.

2.3 The Problem of Getting It Right

Nevermind how we are seen, we will be wise to worry about doing a good job. Here again, it is not just the question of finding true usability problems or making sensible inferences from research, it is also concerned with the ROI for our money and time invested. As Mark Twain Mark Twain once said, “It ain’t what you don’t know that gets you in trouble. It’s what you know for sure that just ain’t so.” Weak usability work runs the risk of steering us towards “knowing” things that ain’t so.

3 What Does Rigor Look Like?

Rigorous means strict, precise, scrupulously accurate. What might this mean, in the context of usability research and practice? It can mean various things. It can mean well-designed experiments with good controls, no confounds, and tests of statistical significance to help us learn about basic human psychological processes. It can mean careful practice, to ensure that tests are performed on representative users (e.g., Huang et al., 2009). It can mean getting empirical about what methods are best, as Molich and his colleagues have in a series of comparative usability evaluations where usability individuals or teams carry out usability evaluations in parallel, (e.g., Molich & Dumas, 2008). It can mean careful attention to potential biases when carrying out qualitative methods. It can mean attending not only to satisfaction data, but also to performance data, including psychological correlates of satisfaction data (e.g., Bias et al., 2010), psychophysical data, and even fMRI data (Huang, 2011). It can mean structured, scholarly preparation for professional practice (e.g., Bias, 2010). And in this session, we’ll hear more about what a rigorous pursuit of usability research and practice might look like.

References

[1] Bias, R.G.: A curriculum for usability professionals (and why we need one). Proceedings of the 3rd International Conference on Applied Human Factors and Ergonomics (AHFE), Miami (July 2010)

[2] Bias, R.G.: Panelist. Discount Testing by Amateurs: Threat or Menace? In: Krug, S. (Chair) Usability Professionals Association Annual Meeting, Baltimore (June 2008)

[3] Bias, R.G.: The dangers of amateur usability engineering. In: Hirsch, S. (chair) Usability in practice: Avoiding pitfalls and seizing opportunities. Annual meeting of the American Society of Information Science and Technology, Long Beach (October 2003)

[4] Bias, R.G., Larson, K., Huang, S.C., Aumer-Ryan, P., Montesclaros, C.: An exploratory study of visual and psychological correlates of preference on onscreen subpixel-rendered text. Journal of the American Society of Information Studies and Technology 61(4), 745–757 (2010)

[5] Bias, R.G., Mayhew, D.J.: Summary: A seat at the table. In: Bias, R.G., Mayhew, D.J. (eds.) Cost-Justifying Usability. Academic Press, Boston (1994)
[6] Huang, S.-C.: Icons: Pictures or logograms? Presentation made at the iConference, Doctoral Colloquium, Seattle (February 2011)
[7] Huang, S.C., Bias, R.G., Payne, T.L., Rogers, J.B.: Remote usability testing: A practice. Poster Presented at the Annual Meeting of the ACM & IEEE Joint Conference of Digital Libraries, June, Austin, TX (2009)
[8] Kortum, P., Bias, R.G., Knott, B.A., Bushey, R.G.: The effect of choice and announcement duration on the estimation of telephone hold time. International Journal of Technology and Human Interaction 4, 29–53 (2008)
[9] Krug, S.: Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems. New Riders, Berkeley (2010)
[10] Krug, S.: Don't make me think: A common sense approach to the web, 2nd edn. New Riders, Thousand Oaks (2005)
[11] Molich, R., Dumas, J.S.: Comparative usability evaluation (CUE-4). Behaviour & Information Technology 27(3) (2008)
[12] Nielsen, J.: Usability for the masses. Journal of Usability Studies 1(1), 2–3 (2005)

HCI Browser: A Tool for Administration and Data Collection for Studies of Web Search Behaviors

Robert Capra

School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA
rcapra3@unc.edu

Abstract. We describe the HCI Browser, a Mozilla Firefox extension designed to support studies of Web information seeking. The HCI Browser presents configurable tasks to the user, collects browser event data as the user interacts with the browser and Web pages, provides mechanisms to record answers that are found, and administers pre- and post-task questionnaires. In this paper, aspects of using and configuring the HCI Browser are summarized and details are given about the events logged, the format of the log files, and how the system is implemented as a Firefox extension. The HCI Browser is open-source software and is available for download at: http://ils.unc.edu/hcibrowser

Keywords: Web information seeking, user interface event logging, data collection.

1 Introduction

Laboratory studies of Web information seeking often involve presenting tasks to participants and recording their actions and behaviors as they use a Web browser to search for information requested by the tasks [4]. Researchers have built and used a variety of tools to help support data collection for Web search and information seeking tasks [1][6][7][8][9][10][11][13][14] and have noted the challenges involved with capturing naturalistic user behaviors [10]. We reviewed a number of approaches [4] and found that they did not meet the needs we had for administering laboratory studies. Specifically, we needed a tool that would: 1) integrate with an existing Web browser to provide a familiar browsing experience to participants, 2) record a wide variety of user interactions with the Web pages and the browser interface, and 3) provide support for administrative aspects of conducting a laboratory study (e.g. administration of questionnaires, closing windows from completed tasks) [4]. To support these needs, we developed the HCI Browser, drawing upon our previous experience building an instrumented web browser [6] and also utilizing concepts from the open-source Lemur Query Log Toolbar project [11]. We have described aspects of the HCI Browser elsewhere [3][4][5]. In this paper, we summarize the functions and operation of the HCI Browser (similar to aspects discussed in [3][4][5]), and extend our prior work by describing in detail the events and data that are logged. We

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 259–268, 2011.

also provide detailed log file examples with explanations and include information about how the HCI Browser is implemented.

The HCI Browser is released as open-source, “alpha” level software. As such, it may contain bugs and we remind potential users to conduct their own testing. Since it is released as open-source software, others can make modifications and improvements to the code.

2 HCI Browser Walk-Through

The HCI Browser is installed as an extension to the Mozilla Firefox Web browser. Because the HCI Browser modifies the Firefox interface and displays pop-up windows and dialog boxes every time Firefox is run, it is recommended to install the HCI Browser into a separate Firefox user profile that is used only for data collection studies. The HCI Browser is designed to: present experimental tasks to users, collect and log browser event data, automatically manage the opening and closing of windows as needed, and to administer optional pre- and post-task questionnaires [4]. In this section, we present the basic functions of the HCI Browser. In the following section, we describe configuration options.

2.1 Start-Up

After installing the HCI Browser extension, every time Firefox is started, a dialog box similar to Fig. 1(a) will be displayed. This dialog box allows the experimenter to enter information that will be used to label data that is logged for the session, including a session number, participant id, and starting task number. Clicking “OK” will advance to a configurable introduction screen, shown in Fig. 1(b), which can be used to display information and instructions to study participants. Clicking “OK” on the introduction screen will advance to an optional pre-task questionnaire screen, or will go directly to the first task if no questionnaire has been configured.

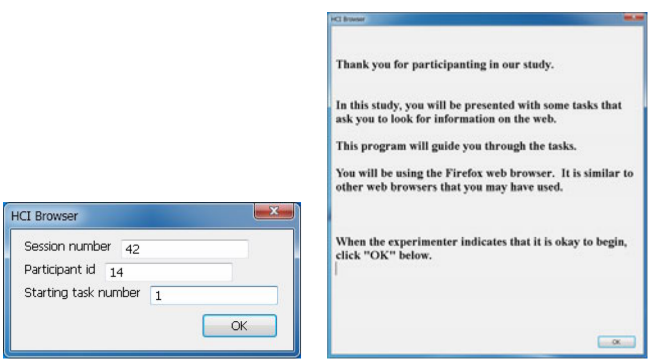


Fig. 1. (a) Start-up dialog box, and (b) Introduction screen

2.2 Pre-task and Post-task Questionnaires (Optional)

The experimenter can design optional pre- and post-task questionnaires to be displayed before and after the participant completes each task. Examples of these are shown in Fig.2 (a) and (b). Three types of questions are supported: multiple choice (responses listed vertically), Likert-type/semantic differential (responses listed horizontally), and open answer (text box). The responses that participants provide are logged and labeled using the information entered by the experimenter on the start-up screen. The task number and prompt text are displayed at the top of the questionnaire screens so that the participant can keep the task context in mind while answering. Note that if configured, the questionnaires are presented before and after **each** task. Different questions can be configured for the pre-task and post-task questionnaires.

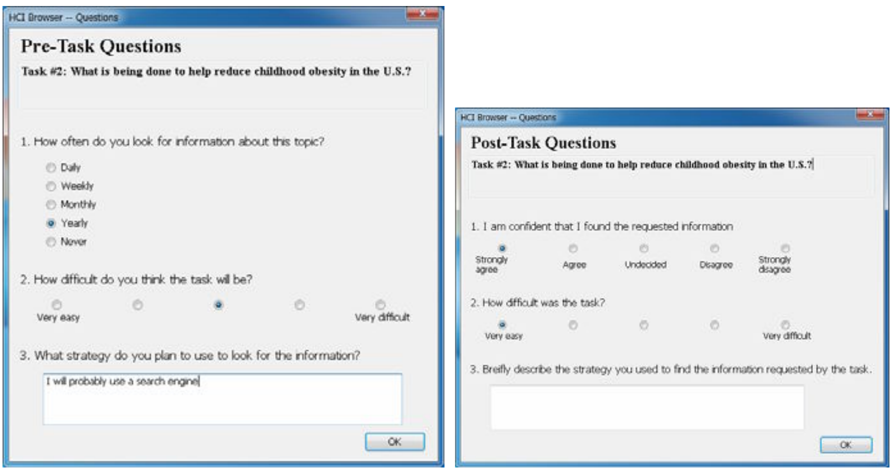


Fig. 2. (a) Pre-task questionnaire, and (b) Post-task questionnaire

2.3 Task Presentation and Answer Collection

After the participant has answered any pre-task questions, the HCI Browser will automatically close the questions window and display a normal Firefox browser window with a toolbar area added above the browser tabs. This section of the window is shown in Fig. 3. In the toolbar area, the task number and task prompt are displayed, along with information about how many answers can be submitted for this task (configurable) and how many answers have been submitted by the user so far. Users can search, browse, and navigate the Web as normal using the browser to look for the information requested by the task. As users work with the browser, their actions are logged as described later in this paper. When they find particular information that they would like to identify as an "answer" for the information requested by the task, they can click the "Found an answer on this page" button in the toolbar. Upon doing so, the right side of the toolbar display changes as shown in Fig. 4 to record the URL of the current web page and to provide a place (labeled "answer text") for the participant to type (or copy and paste) text from the page that is relevant to the task.

When the participant is satisfied with the answer, they can click the “Submit this answer” to continue to look for more answers for the task or they can finish the task by clicking the “Done with this task” button. Tasks can be configured to allow users to submit a maximum number of answers as a method to help set bounds on the amount of time and information requested.

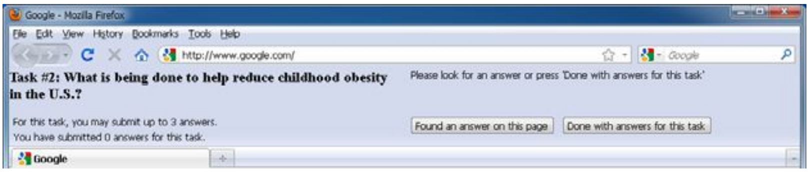


Fig. 3. Toolbar area displaying the task and answer controls

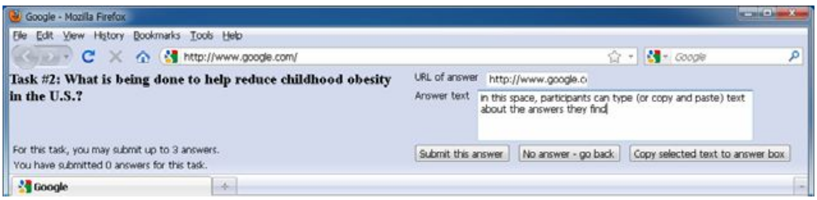


Fig. 4. Toolbar area displaying the answer recording controls

After the participant finishes finding answers for the task (i.e. clicks the “Done with answers for this task” button or submits the maximum number of answers), the HCI Browser will automatically close all open browser windows and display the post-task questionnaire for the task, if it has been configured. If not, then the system will move on to the next task. After completing the last task, the system displays a message letting the participant know that they have completed all the tasks.

3 Configuring the HCI Browser

Configuration of the HCI Browser is accomplished by editing configuration files located in a folder called “*hcibrowser*” on the operating system Desktop. There are four main configuration files: *intro.txt*, *pretask.txt*, *posttask.txt*, and *tasks.txt*. These files are all simple text files that can be edited using any text editor.

The *intro.txt* file contains the text to be displayed on the Introduction screen shown in Fig. 1(b). The *pretask.txt* and *posttask.txt* files are used to configure the question types and question text presented in the pre- and post-task questionnaires respectively. If these either of these files is not present in the “*hcibrowser*” configuration folder on the Desktop, then the respective questionnaire will not be presented to the participants. The format for the *pretask.txt* and *posttask.txt* files is identical. Fig. 5 shows the *pretask.txt* file that corresponds to the questionnaire in Fig. 2(a).

```
01  MultipleChoice
02  5
03  How often do you look for information about this topic?
04  Daily
05  Weekly
06  Monthly
07  Yearly
08  Never
09  ---
10  LikertType
11  5
12  How difficult do you think the task will be?
13  Very easy
14
15
16
17  Very difficult
18  ---
19  OpenAnswer
20  What strategy do you plan to use to look for the information?
21  ---
```

Fig. 5. Example pre-task questionnaire configuration file. Line numbers are shown here for description and are not included in the actual configuration file.

Three types of questions are supported:

- Multiple choice (Fig. 5, lines 01-09) – response choices are displayed vertically. The keyword “MultipleChoice” is used in the configuration file (line 01), followed by a line indicating how many choices there will be (line 02), a line for the question text (line 03), followed by the text of the choices, one choice per line (lines 04-08). A required line of three minus signs (line 09) indicates the end of the question.
- Likert-type/semantic differential (Fig. 5, lines 10-18) – response choices are displayed horizontally. The keyword “LikertType” is used in the configuration file (line 10), followed by a line indicating how many choices there will be (line 11), a line for the question text (line 12), followed by the text of the choices, one choice per line (lines 13-17). In this example, lines 14-16 are blank because only endpoint anchor text was desired for this scale. Line 18 is the required question terminator line.
- Open answer (Fig. 5, lines 19-21) – the question is presented along with a text box for the participant to type an open response. The keyword “OpenAnswer” is used in the configuration file (line 19), followed by a line for the question text (line 20). Line 21 is the required question terminator line of three minus signs.

The final configuration file is named *tasks.txt* and contains information about the tasks to be presented. An example is shown in Fig. 6. Each task to be presented requires two lines in the configuration file: an initial line that contains the text of the task prompt (e.g. lines 01, 03, and 05 in Fig. 6), and a second line that contains the maximum number of answers that will be accepted for the task (e.g. lines 02, 04, and 06). The tasks will be automatically numbered and presented in the order in which they appear in the *tasks.txt* configuration file.

```
How tall is the U.S. capital building in Washington, DC?
1
What is being done to help reduce childhood obesity in the U.S.?
3
What are some reported causes of global warming?
3
```

Fig. 6. Example *tasks.txt* configuration file. Line numbers are shown here only for description and are not included in the actual configuration file.

4 Events and Data Logging

Events and data collected while running the HCI Browser are logged into an automatically created sub-folder called "*data*" in the "*hcibrowser*" folder on the Desktop. Within the "*data*" folder, additional sub-folders are created for each session and participant id entered at the start-up screen. Individual text-based log files are placed in these folders for each task and are named based on the session, participant id, task number, and time stamp (e.g. S42_P14_T2_1297118649034.txt). The log files contain data from the pre- and post-task questionnaires (if administered) and from the participant's interactions with the browser during the task.

Events logging is based on the HCI Browser observing Firefox browser events, filtering and interpreting these events, and writing them to a log file. Many of the events used by the HCI Browser are similar to underlying Firefox events [12] and also to event labels used by the Lemur Query Log Toolbar project [11]. Table 1 lists a subset of common events that are logged. Additional events are logged for actions that involve the browser history and bookmark menus.

5 Understanding the Log Files

The log files are organized as a text file with one event per line. Information on each line is separated by tab characters (i.e. tab delimited), making it easy to read the log file into a spreadsheet or other analysis program. Fig.7 shows an example line of a log file. The first eight fields on each line are fixed and contain information as follows. The first field is a date+time stamp in numeric format (1297118695677), followed by a human-readable date (7-2-2011 = Feb 7, 2011), and time (17:44:55 = 5:44pm and 55 seconds).

The next three fields are the session number (S42), participant id (P14), and the task number (T2). The next field indicates if the current line was logged during the pre- or post-task questionnaires ("pretask" and "posttask"), or while the participant was doing the task itself ("intask"). The final fixed field indicates the event and uses codes as outlined in Table 1.

Table 1. Events logged by the HCI Browser

Event	Event code	Description
Page load	LoadCap	A page is loaded into a window or tab. This event does not mean that the page is being viewed (i.e. in focus). Based on the Firefox "load" event.
Focus	Focus	The focus has changed, meaning that the user is now viewing a different tab or window than they were before the focus change. Based on the Firefox "focus" event.
Scrolling	Scroll	Scrolling on the page. This event is logged based on a timer such that subsequent scrolling is not logged for 500 milliseconds.
Left click, right click	LClick, RClick	The user has clicked the mouse on a page element. Left-clicks typically are link clicks. Right-clicks often are to open a link in a new tab or window via the right-click context menu.
Add, select, remove tab	AddTab, SelTab, RmTab	Actions dealing with browser tabs. Also logs a set of informational messages listing currently open browser windows and tabs. SelTab typically also results in a focus change event.
Information	Info	Informational messages (e.g. list of open tabs)
Typing URLs	TypeURL	The user has typed into the browser's address bar (e.g. typed in a URL to navigate to it)
Browser button clicks	[Home|Back|Forward|Reload] ButtonClick	User clicked the browser's home button, back button, forward button, or reload button.

.

```
1297118695677  7-2-2011  17:44:55  S42  P14  T2  intask  LoadCap
http://www.cnn.com/  http://www.cnn.com/
```

Fig. 7. Example line of a log file

After the eight fixed fields, the remaining logging information on each line depends on the event. Many events record the URL of the page that was in focus when the event occurred, and/or the URL of the page that is the target of the event. Events also include a final field that contains the most recently focused URL (which may be the same as the other URL listed).

Fig. 8 shows a portion of a log file that illustrates logging of the interactions and answers given by a user for a pre-task questionnaire. Note that in Fig. 8 parts of the log file have been omitted for space. When the pre-task questions are displayed, information is logged about the task description (e.g. at time 7110 in Fig. 8) and the maximum number of answers for this task (7113). At time 7732, the user finished entering answers to the pre-task questions and clicked "OK" to move on to start the task. The answers provided to questions q1, q2, and q3 are then logged (7736, 7739, 7741) and at time 7893 the log switches to logging events "intask".

```
TIME    EVENT     DETAILS
7110    pretask   TaskDescription  What is being done to help
7113    pretask   TaskMaxAnswers      3       null
7732    pretask   onAccept event -- ok pressed for Task #2
7736    pretask   q1   4       null
7739    pretask   q2   3       null
7741    pretask   q3   i plan to use search engines and look
7893    intask    Focus about:blank    about:blank
```

Fig. 8. Example pretask section of a log file

Fig. 9 shows a more extensive example of a log file, again with portions omitted for space. This section shows that at time 5677, the user loaded a Google search results page and then right-clicked on a link (0323). The right-click was to load the link in a new tab (1798). When a new tab is opened, the HCI Browser logs a summary of all the open windows and tabs (1821, 1824, 1828). At the time the summary was recorded (1828), tab #1 had not yet loaded the new page, so it is listed as url=about:blank, but the load event for the www.cdc.gov page is next in the log (2865). Notice that the www.cdc.gov page is loaded in the new tab but the user does not focus away from the Google search results page at this point. The user is still interacting with the Google search results and issues several scroll events to browse the results (5384, 6539, 7174). The user then right-clicks and loads two other pages from the search results (a page from aspe.hhs.gov and an en.wikipedia.org page), again without navigating away from the Google search results page. Finally, at 3823, the user selects a different tab, switching the focus to the www.cdc.gov page (3844). The user then scrolls on this page (5470, 6271), and left-clicks a link on the cdc.gov site about “keystrategies” (9875).

6 Implementation

The HCI Browser is written as a Firefox Web browser extension, packaged as an XPI file. The browser interface components are written in the XUL user-interface description language and the program logic is implemented in Javascript. It is outside the scope of this paper to provide details about Firefox extension development, but there are many good tutorials (e.g. [2]). The concepts implemented in the open-source Lemur Query Log Toolbar are also helpful in understanding Firefox toolbar creation and event logging and the HCI Browser has been influenced and informed by many of its concepts [11].

The HCI Browser monitors three main types of events in the browser: 1) events in the browser user interface (e.g. back/reload button, typing in the address bar, menu activity), 2) events that deal with windows, tabs, and page loading, and 3) events that occur on the currently focused page (e.g. mouse clicks). In order to monitor the mouse events on the Web pages, when each page is loaded, a “mousedown” event handler is added to every link found on the page.

Sometimes a single Web page will cause several underlying Firefox “load” or “focus” events to occur. The HCI Browser employs filters so that if multiple, sequential LoadCap or Focus events are detected for the same URL, only the first will be written to the log file.

```
TIME    EVENT     DETAILS
5677    LoadCap   http://www.google.com/search?hl=en&source=hp&b
0323    RClick    http://www.google.com/url?sa=t&source=web&cd=1
1798    AddTab    about:blank  http://www.google.com/search?hl=en
1821    Info      Currently open windows and tabs are logged bel
1824    Info      Window=0, tab=0, url=http://www.google.com/sea
1828    Info      Window=0, tab=1, url=about:blank   http://www
2865    LoadCap   http://www.cdc.gov/HealthyYouth/obesity/   ht
5384    Scroll    http://www.google.com/search?hl=en&source=hp&b
6539    Scroll    http://www.google.com/search?hl=en&source=hp&b
7174    Scroll    http://www.google.com/search?hl=en&source=hp&b
0043    RClick    http://aspe.hhs.gov/health/reports/child_obesi
0996    AddTab    about:blank  http://www.google.com/search?hl=en
1009    Info      Currently open windows and tabs are logged bel
1011    Info      Window=0, tab=0, url=http://www.google.com/sea
1032    Info      Window=0, tab=1, url=http://www.cdc.gov/Health
1050    Info      Window=0, tab=2, url=about:blank   http://www
1946    LoadCap   http://aspe.hhs.gov/health/reports/child_obesi
3033    RClick    http://www.google.com/url?sa=t&source=web&cd=1
4101    AddTab    about:blank  http://www.google.com/search?hl=en
5004    Scroll    http://www.google.com/search?hl=en&source=hp&b
6256    LoadCap   http://en.wikipedia.org/wiki/Childhood_obesity
3823    SelTab    http://www.cdc.gov/HealthyYouth/obesity/   ht
3844    Focus     http://www.cdc.gov/HealthyYouth/obesity/   ht
5470    Scroll    http://www.cdc.gov/HealthyYouth/obesity/   ht
6271    Scroll    http://www.cdc.gov/HealthyYouth/obesity/   ht
9875    LClick    http://www.cdc.gov/HealthyYouth/keystrategies/
```

Fig. 9. Example log file. Parts of the file have been omitted and the time stamp has been shorted in order to fit the space.

7 Summary

The HCI Browser is a tool to help support data collection and administration of studies of Web information seeking. It is released as "alpha" level, open-source software. More details, including information about downloading and installing the HCI Browser are at: http://ils.unc.edu/hcibrowser

References

1. Bierig, R., Gwizdka, J., Cole, M.J.: A user-centered experiment and logging framework for interactive information. In: Proceedings of the SIGIR 2009 Workshop on Understanding the User: Logging and Interpreting User Interactions in Information Search and Retrieval, Boston, MA, July 23, pp. 8–11 (2009)
2. Bishop, J.: Born Geek Firefox Toolbar Tutorial, http://www.borngeek.com/firefox/toolbar-tutorial/ (retrieved February 10, 2011)
3. Capra, R.: HCI Browser: A Tool for Studying Web Search Behavior. Poster Presentation at the 73rd American Society for Information Science and Technology Annual Meeting (2010)

4. Capra, R.: The HCI Browser Tool for Studying Web Search Behavior. In: Proceedings of the Third Workshop on Human-Computer Interaction and Information Retrieval, pp. 38–41 (2009)
5. Capra, R.: HCI Browser: A Tool for Studying Web Search Behavior. In: Workshop on Understanding the User at ACM SIGIR (2009)
6. Capra, R.: Studying Elapsed Time and Task Factors in Re-Finding Electronic Information. In: Workshop on Personal Information Management at ACM CHI (2008)
7. Claypool, M., Le, P., Wased, M., Brown, D.: Implicit Interest Indicators. In: Proceedings of the 6th International Conference on Intelligent User Interfaces (2001)
8. Feild, H., Allan, J., Jones, R.: Predicting Searcher Frustration. In: The Proceedings of the 33rd International ACM Conference on Research and Development in Information Retrieval (SIGIR 2010), Geneva, Switzerland, July 19-23 (2010)
9. Kaasten, S. and Greenberg, S.: URL Tracer: URL Tracing Software for Internet Explorer, `http://grouplab.cpsc.ucalgary.ca/cookbook/index.php/Utilities/URLTracer` (retrieved on February 7, 2011)
10. Keller, M., Hawkey, K., Inkpen, K., Watters, C.: Challenges of Capturing Natural Web-Based User Behaviors. International Journal of Human-Computer Interaction 24(4), 385–409 (2008)
11. Lemur Query Log Toolbar, `http://www.lemurproject.org/querylogtoolbar`
12. Mozilla Developer Network Events, `https://developer.mozilla.org/en/XUL/Events` (retrieved February 10, 2011)
13. Reeder, R.W., Pirolli, P., Card, S.K.: WebEyeMapper and WebLogger: Tools for Analyzing Eye Tracking Data Collected in Web-Use Studies. In: CHI 2001 Extended Abstracts (2001)
14. Turnbull, D. Webtacker: A Tool for Understanding Web Use, `http://www.ischool.utexas.edu/~donturn/research/webtracker/index.html` (retreived on February 7, 2011)

Design and Evaluation of the Customized Product Color Combination Interfaces Using 3D Model and 2D Illustration Display

Cheih-Ying Chen[1] and Ying-Jye Lee[2,*]

[1] Department of Cultural and Creative Industries, National Pingtung University of Education, 4-18 Minsheng Road, Pingtung City, Pingtung County 90003, Taiwan, R.O.C.
[2] Department of Cultural Business Development, National Kaohsiung University of Applied Sciences, 415 Chien Kung Road, Kaohsiung 807, Taiwan, R.O.C.
Tel.: +886-7-3814526 ext.3249
yjlee@cc.kuas.edu.tw

Abstract. The objective of this study is to investigate the interactive relationship between product color information and color combination interface on computer screen. In order to achieve the objective, this study takes the cell phone as an example. Also, two customized product color interfaces based on both three dimensional model product model and two dimensional product illustration displays via the marketing approach of experience economy are designed in the study. Furthermore, this study discusses user interface satisfaction of customized product color combination selection. It shows that both the interactive process and the resulting differ in three dimensional model display and two dimensional illustration display. It seems to be the best way for users to get a unique experience and a realistic feeling of the virtual product in 360 degrees with three dimensional model product model for displaying customized product color combination.

Keywords: customization, cell phone, image compositing, interface design, 3D model.

1 Introduction

In the new type service and experience with novel ideas involving consumers on the interaction to lead them feel, like and change service into unforgotten experience. Consumers are affected by the experience economy with either virtual or real environment more than the traditional marketing (Pine & Gilmore, 1999). Customization is one kind of experience economy tools. The diversified and customized products gradually substitute for the unified and standardized products, and then apply them to the e-commerce. Also, numerous products use modularity changes and mass customization to create many product variations and styles. For

* Corresponding author.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 269–275, 2011.

example, modularization of an automobile allows buyers to customize features such as exterior color, engine power, interior color and safety devices (Salvador, Forza & Rungtusanatham, 2002). More companies adopt customization for consumers undergoing an experience of design themselves to enhance purchase willingly. Wu et al. (2005) noted different color products can achieve different visual effects and create more pleasing and stylish product image. Therefore, different color combinations by applying image compositing technique to the customized design interface; it will give the consumers different color fascination and design experience (Lee et al. 2009). In the cell phone industry, there are many choices in colors and patterns for covers to change the look of cell phone completely. (e.g., Sony Ericsson, www.styleupcover.com: Design your own gadgets). There are many choices in devices and colors in automobile industry; "Smart" automobile by DaimlerChrysler Mercedes-Benz is particularly well-known for its mass customization in color choices (www.smartusa.com/smartexpressions).

2 Three Dimensional Product Model and Two Dimensional Product Illustration

As Three dimensional models reach broader acceptance, their use as communication media is attracting both commercial and industrial interests. Some companies have gradually adopted three dimensional models in place of product photographic in commercial web. Three dimensional models also in place of two dimensional illustrations for displaying image compositing in the few customized design interfaces. So far, little is known about user requirements and cognitive aspects of three dimensional model display for product customization. Also the latter's integration into the virtual product development is far from being discussed. The objective of this study is to investigate the interactive relationship between product color information and color combination interface on computer screen. In order to achieve the objective, this study takes the cell phone as an example. Also, two customized product color interfaces based on both three dimensional product model and two dimensional product illustration displays via the marketing approach of experience economy are designed in the study. Furthermore, this study will discuss the interaction of the consumer on the interface.

3 Methods

3.1 Interface Arrangement

An idealized mobile phone was divided into three product modular sections: top cover, base and camera face plate. Mobile phones are generally monochrome, bicolor or tricolor. In this study, the five common colors (white, silver, black, blue and red) were offered for top cover and camera face plate. Black was offered for the base. The study refer to the layout type of itemized color chips, which is divided into some items according to the customized color module parts of product(Wu et al. 2010), has the best grouping type for customized product color combination selection in search time by users (Chen et al. 2007). Itemized color chip by modular section, disposed the color sample displays by product color modular sections (5 monochrome chips * 5 monochrome chips * 1 monochrome chips =25). Three types of color combination were available: monochrome (one color throughout: 1 combinations), bicolor (two colors, one repeated: 12 combinations) and tricolor (three colors, none repeated: 12 combinations) for a total of 25 combinations.

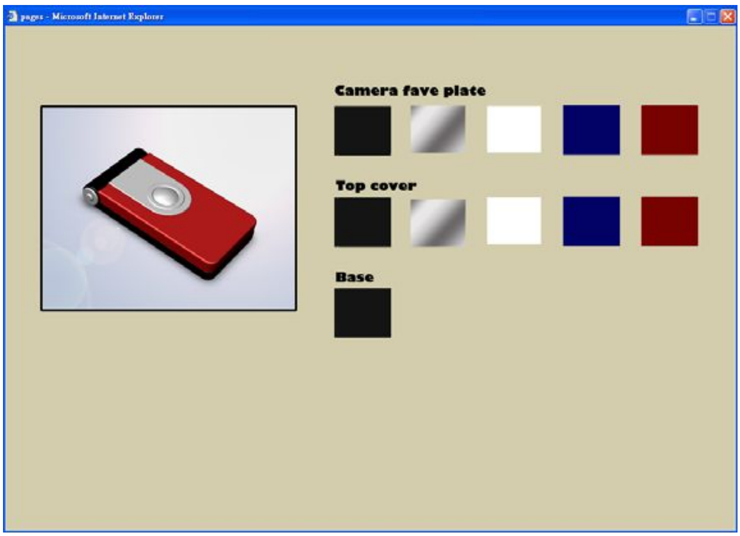


Fig. 1. The interface of two dimensional static illustration for customized product color combination selection

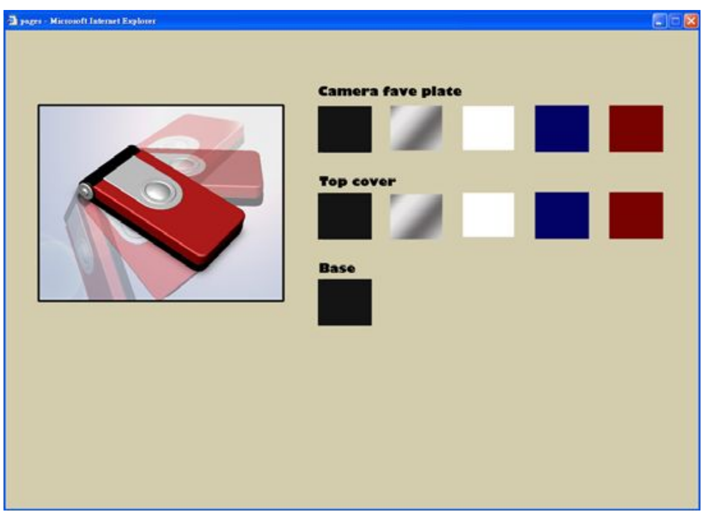


Fig. 2. The interface of three dimensional dynamic model for customized product color combination selection

In this experiment, these color chips were also nodes to link the corresponding larger compositing image of on the left side of a computer screen. The larger compositing image display of mobile phone color combination was assigned 2 types: two dimensional static illustration (Fig.1) and three dimensional dynamic model (Fig.2).

3.2 Participants

The 30 participants in the experiment (survey of layout and layout element preference) were paid from various departments of the college. All were between the ages of 19 and 26 (Mean age=21.3; SD=2.41). All participants had a color vision test found none to be color blind.

3.3 Apparatus

The experimental stimuli (screen layouts) were displayed by a multimedia computer (CPU: AMD Athlon 2.0GHz; RAM: 512MB PC3200 DDR) on a 19-inch monitor with 1280X1024 resolution and 75HZ refresh rate.

3.4 Procedure

Each subject played with each of the 2 interfaces and completed a uniform set of three color combination selection tasks:

1. Monochrome: black top cover, base and camera face plate
2. Bi-color: white top cover, black base, white camera face plate.
3. Tri-color: red top cover, black base, silver camera face plate.
4. Fine your favorite color combination.

Each subject played with each of the 2 interfaces, then filled questionnaire for User Interface Satisfaction for each interface. This study adopted a measure of overall system satisfaction along six scales of the questionnaire for User Interface Satisfaction (QUIS) on a 9-point scale (Chin, Diehl, and Norman, 1988).

4 Results and Discussion

Mean of overall user interface satisfaction for all participants in the color combination selection task is given in Table 1.

Table 1. Mean of overall user interface satisfaction of all participants

Independent variable	Two dimensional static illustration		Three dimensional dynamic model		p-value
	Mean	S.D.	Mean	S.D.	
Terrible / Wonderful	6.07	0.94	7.43	1.01	0.000000
Difficult / Easy	7.17	0.95	7.53	0.97	0.025108
Frustrating / Satisfying	6.03	1.07	7.93	0.74	0.000000
Inadequate power / Adequate power	5.27	0.94	5.90	0.71	0.000092
Dull / Stimulating	5.30	1.06	7.53	1.01	0 .000000
Rigid / Flexible	6.53	0.78	7.00	0.87	0.013638

4.1 Terrible / Wonderful

T-test analysis of the wonderful item find the results of product displays of two interfaces to be significant (p=0.000000 < 0.01), indicating that two interfaces significantly affected wonderful interaction.

Results of t- tests are given in Table 1. On the wonderful item, three dimensional dynamic model (mean score=7.43) is significantly superior to two dimensional static illustration (mean score=6.07). Therefore, product display of three dimensional dynamic model is generally preferred, because users could have a unique experience with interaction.

4.2 Difficult / Easy

There was no significant difference in two interfaces with two dimensional static illustration (mean score=7.17) and three dimensional dynamic mode (mean score=7.53). In fact, users considered the two interfaces easy to use and give both high scores.

4.3 Frustrating / Satisfying

T-test shows the results for satisfying (p=0.000000< 0.01) to be significant. Product display significantly affect mean scores on the satisfying item between two interfaces. Three dimensional dynamic model (mean score=7.93) achieves significantly higher mean scores on satisfying item than two dimensional static illustration (mean score=6.03). Therefore, product display of three dimensional dynamic model is generally satisfied, because users could see the color of every part of the product clearly.

4.4 Inadequate Power / Adequate Power

T-test shows significant differences on the adequate power item between two interfaces (p=0.000092<0.01). Results of t-tests show that mean scores for three dimensional dynamic model (mean score=5.90) is significantly higher than two dimensional static illustration (mean score=5.27). Therefore, product display of three dimensional dynamic model is generally more powerful, and could improve users' recognition of spatiality and spatial views.

4.5 Dull / Stimulating

T-test shows significant differences on the stimulating item between two interfaces (p=0.000000< 0.01). Results of t-tests show that mean scores for three dimensional dynamic model (mean score=7.53) is significantly higher than two dimensional static illustration (mean score=5.30). Therefore, product display of three dimensional dynamic model is generally stimulating, because users could view the animation and play with interface of three dimensional dynamic model display.

4.6 Rigid / Flexible

There was no significant difference on the flexible item between two interfaces with two dimensional static illustration (mean score=7.00) and three dimensional dynamic mode (mean score=6.53). In fact, users considered the two interfaces are flexible and give both high scores.

5 Conclusion

In conclusion, this study primarily uses cell phone color combination to investigate the users' operational preferences in terms of product display with two dimensional static illustration or three dimensional dynamic model for customized color selection. It shows that both the interactive process and the resulting differ in three dimensional model display and two dimensional illustration display, namely in terms of the perceived appropriateness of interface, in terms of the perceived fascination for the display, and in terms of the perceived stimulation by the medium and the aspects of visual view. Results of the experiment can be applied to similar products, and summarized into the following suggestions:

1. Three dimensional dynamic model of color combinations yields wonderful interaction for users. Product display with three dimensional dynamic model for customized color selection is more exciting than two dimensional static illustration, and it gives a strong interest in using the three dimensional model display as a medium for image compositing design and its additional dynamic externalization processes. It seems to be the best way for users to get a unique experience and a realistic feeling of the virtual product.
2. Product display with three dimensional dynamic model is stimulating and satisfied for a clear identification of different product parts and must be facilitated by 360 degree view. Users can view various color combination of product modular sections with two dimensional static illustration or three dimensional dynamic mode. Three dimensional dynamic model provides new views for the customizing process, namely animation and product viewpoint change. With a three dimensional image, a customized product can be spun and viewed in details form various angles in 360 degrees. Viewing and comparing all color combinations are best facilitated by product thumbnails by the three dimensional dynamic model.
3. Product display with three dimensional dynamic model adequates more power than two dimensional static illustration, and interact with users when viewing and comparing all options are interactive with users. The result reveals that the biggest advantage of three dimensional models lies in the process. In particular, participants emphasized the interface's ability to foster product concept and improve the recognition of spatiality and spatial views.
4. Both three dimensional model display and two dimensional illustration display for customized color selection are well interface design for interaction with users. As users progress towards three dimensional models, new experience interactive interfaces are needed, which account for the user's perceptual and cognitive abilities for the customization and experience.

References

1. Chen, C.Y., Lee, Y.J., Wu, F.G., Su, C.F.: Screen Layout on Color Search Task for Customized Product Color Combination Selection. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551, pp. 32–40. Springer, Heidelberg (2007)
2. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In: ACM CHI 1988 Conference on Human Factors in Computing Systems, pp. 213–218 (1988)
3. Wu, F.G., Chen, C.H., Lee, Y.J., Chen, R.: Effects of color sample display and color sample grouping on screen layout usability for customized product color selection. Computers in Human Behavior 26, 51–60 (2010)
4. Lee, Y.J., Chen, C.H., Wu, F.G.: Design and evaluation of the customized product color combination interface based on scenario experience, HCI 2009. LNCS, vol. 5617, pp. 263–2700. Springer, Heidelberg (2009)
5. Joseph Pine II, B., Gilmore, J.H.: The Experience Economy: Work Is Theatre & Every Business a Stage. Harvard Business School Press, Boston (1999)
6. Salvador, F., Forza, C., Rungtusanatham, M.: Modularity, product variety, production volume, and component source theorizing, beyond generic prescriptions. Journal of Operations Management 20, 549–575 (2002)
7. Wu, F.G., Chen, C.H., Lee, Y.J.: Optimum product color combination using the FAHP approach combined with the image compositing technique. In: HCI International 2005, Las Vegas, USA (2005)
8. Zaichkowsky, J.K.: Measuring the Involvement Constructure. Journal of Consumer Research 12(3), 341–352 (1985)

The Inmates Are Still Running the Asylum: How to Share a Design Vision with Engineers

Uday Gajendar and Colin Johnson

Citrix Systems, Inc. 4988 Great America Parkway
Santa Clara, CA 95054 USA

Abstract. This paper presents an approach taken by Citrix to shape a balanced, shared product design effort with engineering. Key points include the rise of hybrid designers skilled in software programming, the use of standard UI components, and collaborative standards council activities. Action items are also noted for interested readers trying to build their own integrated design and development efforts for good software user interfaces.

Keywords: User interface design, user interface engineering, user interface technology.

1 Introduction

In his seminal 1999 book *The Inmates are running the asylum*, Alan Cooper made a bold claim that shook the software industry. “The reason for bad software products,” he argued, “is that engineers are in control.” [1] Because engineers tend to craft interfaces the way they craft code—logically and mechanically—software was invariably designed *by* engineers, *for* engineers, who gave little thought to users, and what inspires and delights them.

Much has improved since the book’s publication. Interaction designers, visual designers, and researchers are increasingly involved in developing software products at many companies. Businesses have noticed the commercial successes of Apple, Salesforce, Netflix, and the like – companies that feature design, and have design-driven product development processes. A cursory look at recent IxDA or CHI job postings reveals an increased demand for hiring designers. At conferences and in online discussion, user-centered, design-driven product-development approaches are more and more widely accepted.

But there is still a long way to go, partly because of the rapid spread of user interface (UI) technologies that enable a range of outcomes. Software development teams can now choose from a wide variety of technologies to use in building application interfaces. Each technology is good at enabling certain sorts of use experiences, on various devices, while doing specific tasks. For the most part, these decisions are owned by engineering teams, who, all too often, make choices for reasons that have nothing to do with giving users a good experience. Instead, they select a UI technology because it enables them to showcase “cool” features, because they want to learn to use it, or because time is short and they already know how to use

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 276–282, 2011.

it. The result, too often, is that user experience is an afterthought, and suffers accordingly.

This is particularly true when, as too frequently happens, engineering teams switch UI technologies from project to project. Different technologies enable different sorts of UI behaviors, and different looks. Designers have to scramble to keep up. Often they have to migrate branded assets, and other standard UI elements, to a new, unfamiliar technology framework (from HTML to WPF, for example, or Silverlight to Objective C). Worse, designers sometimes have to re-design everything from scratch, consuming valuable time and resources due to color bit-depth differences or font support limits.

So not only has Cooper's problem not been solved—in some ways, it has become worse. As Cooper points out, engineers are smart, pragmatic, and territorial. It is one thing to persuade an engineer that she should not design a graphical user interface. It's quite another to persuade her that she should not choose the technology used to build a certain interface. Yet this is what software companies must do, to avoid the sorts of problems we have just outlined.

Rather, these and related decisions should be vested in a new sort of product team, one in which developers, designers, and business-side employees collaborate at every step. Moreover, these designers should have strong coding skills. Designers who can code can make informed contributions to team technology decisions, and produce not just static wireframes, but fully coded prototypes that can serve as the basis for development of final products.

This paper will showcase our efforts, at Citrix, to build product teams of this sort. It will also discuss the effects of our decision to staff the Citrix Product Design group with designers who have the hybrid skills needed to partner fully with engineers, to create products that are both technologically advanced, and deliver a top-notch user experience that is consistent across product lines.

2 Explosion of UI Technologies

Fifteen years ago the UI technology choice was simple. Write a Windows application or a simple HTML-based interface for the web browser. Now, product teams have a dizzying array of choices. Consider the list on this Wikipedia page, which shows the variety of frameworks available for building web-based applications. There are also myriad component toolkits available that are independent of the underlying application framework and commonly run on top of them. Examples include ExtGWT or SmartGWT on top of GWT, jQuery or Prototype on top of Ruby on Rails, and Infragistics or Telerik on top of WPF Silverlight.

Each such technology makes it relatively easy to enable a particular look and feel and behaviors, and difficult to enable others. Unfortunately, at many software companies, such as Oracle, Cisco, and Adobe, HCI-trained designers are rarely or never involved in choosing the UI technology. As a result, they have little or no impact on a decision that radically constrains their ability to ensure that their designs, and their user experience vision, are realized in the final product.

3 Towards a New Kind of Product Design Team

At Citrix, the Product Design group currently supports a relatively narrow range of products, with most intended to be used by IT personnel. And most are used on Windows-based desktops. But this is changing. We are building more and more applications for consumer users, and for use on a variety of OS and device platforms, including smartphones and tablets.

To make sure these applications deliver a high-quality experience that is consistent across OSs and devices, it's essential for our group to cooperate with Citrix's engineering team at every step. A critical part of this is playing a role in selecting the appropriate UI technology to be used in each product. How do we do this?

3.1 Ensuring an Optimal UI via Components and Standards

Our group has set up a team that is developing, and making available to product teams, pre-themed branded instantiations of common UI components, for a select set of UI technologies and platforms. Previously, engineering teams had to implement designs from wireframes and specs created by designers. Under the new system, engineers can create products from pre-built components. This relieves both them and designers of much of the grunt work involved in crafting a UI.

This effort would not be worthwhile, however, if each component were not designed and built to be optimally functional, deliver high-quality interactions, and look, feel, and act like other Citrix applications. Fortunately, our component-production process ensures that this is the case.

Components team composition is a key to the success of this effort. Critically, Product Design UI thought leaders are involved in every step of the component creation process. A designer is of course involved every step of the way to ensure total user experience quality. Full-time team members include a Product Design Group UI architect, who sets the vision and direction and initially acts as the development lead for components. The architect writes and maintains the architecture specification for each component, and writes and reviews functional and design specifications. The architect also creates components and collaborates with part-time development leads in defining common component architecture, and defining services that work with the components. There are full-time component developers who work on creating components and full-time QA personnel who test them.

The team also includes a project manager and a product manager. The project manager ensures that everything runs smoothly, facilitates discussions between teams, and so forth. The project manager also maintains an up to date list of all components and services currently supported and a roadmap for future work with estimated "drop" dates. The product manager writes and maintains the Product Requirements Document (PRD) setting the overall requirements the component library needs to meet. The product manager also fields feature requests for custom components and feeds them in as requirements.

On the part-time side, UI developers float in and out on an as-needed and as-available basis. They work on component creation as well as specification and code reviews. Others can get involved as they have time and the inclination, because the components effort is run like an open source project. They can write and test components, help with documentation, and build sample applications. If the components library is lacking a particular component, and the components team can't build it in time to be included in a particular product release, the product team is encouraged to write it and later lobby to have it included in the library.

3.2 A Demo Car Built from UI Components

We need not just build components, but ensure their adoption by product teams, and give guidance on how they can and should be used. One way we're doing this is by building and publicizing a "demo car" – a sample application whose purpose is to showcase a set of complex yet widely useful components from the component library. It will have a generic look and feel in line with the desired look and feel of all our applications. All components will be used in the intended fashion and display archetypal interactions. Accompanying the "demo car" will be, for each component, a set of Functional Specifications that detail the public API, and the associated unit tests, and a set of Design Specifications that detail how the component is put together internally. Eventually, the parent UI pattern behind each component will be housed in a linked pattern library. Each pattern definition will include a link to the component, as well as code samples showing how to instantiate and use it.

3.3 Team Activities

Other activities encourage other groups across Citrix to partner with the Product Design group, and ensure the user-experience considerations shape product development at every turn.

Design Process Roadshow. The authors and other design directors recently traveled to Citrix branch offices in Cambridge UK, Bangalore, and Sydney, to advocate in person for the use of common UI components and design patterns, and give updates on the progress of efforts to develop both for use in building applications.

They also educated teams on Citrix's newly rolled out UCD Design Process, which mandates that user needs be clearly identified and taken into full account at every stage, and ensures close, continual collaboration between designers and engineers. Key to both is the mandate that each for each product, a detailed "design brief" be developed and adhered to. Each such brief captures technical as well as user goals. Throughout the design process, it serves as a means of ensuring that both sets of goals are met by the product design. It also helps prevent confusion, later in the course of product development, about what exactly is to be built and what tools are to be used.

Standards Council Meetings. The new Standards Council exists to define and approve standards and standard UI patterns and components. It also disseminates standards, particularly to external partners, including product managers, QA and customer support personnel, and technical documentation leaders. Emphasis is placedupon "soft enforcement" through partnerships and constructive reviews, not adamant policing with strictly defined scorecards. Any product team resistance is channeled up through our executive support chain, including the VP of Product Design. Weekly Standards Council meetings also serves as a valuable opportunity to survey what is working and what isn't, from a standards perspective, to review new

UI requests, communicate UI updates to teams, discuss revised requirements, and so forth. In this way, the Standards Council ensures that standards and components will be used in all UI design and development work, across Citrix.

Tiger Teams. These are temporary, dedicated teams of visual & UI designers, each of which resolves tactical UI implementation issues for products of a similar genre (administrative monitoring consoles, for example). Each team also works with the Standards Council to help the Council extend and revise UI standards to cover cases of the sort it works on. As UI guidelines become more complete, mature, and propagated, the Tigers Teams will work only on emergencies, at executive request.

4 Towards a New Kind of Designer

Our designers do not rely solely on management's help to make this new system work. They also know how to work constructively with engineers, to ensure that their relationship is one of useful collaboration, and meaningful understanding. Particularly valuable here is their level of technical knowledge – at Citrix, we prioritize hiring designers who can code.

We believe that software designers can no longer get by on their right-brain skills alone. They need not just to mock up attractive interfaces, but also code well enough to build interactive prototypes that show exactly how an application should work. They also need a deep understanding of engineering practices, in order to understand the engineering implications of the design choices they make. By developing these skills, they'll be able to more productively collaborate with engineers, both when working in the sort of cross-disciplinary team we discuss above, and in traditional software environments.

Specifically a hybrid designer should be able to:

- Critically evaluate UI technologies and discuss their merits with engineers
- Translate interface layouts into functional prototypes with clean, modular scripts and code snippets that can be reused for subsequent builds
- Effectively and constructively call out an engineer when that person resists design-driven change to a product or process
- Be able to argue back effectively and generate workaround design alternatives that map to the engineer's technical concerns.

Ultimately HCI professionals need to be more aware of the **capabilities** of a chosen technology and how this affects their design. For instance:

- Can the chosen technology be re-skinned and styled to get the desired design?
- Can it be made responsive enough in the environment we intend to use it?
- When the engineering team says it cannot be done, do you believe them? Are you able to suggest an feasible alternative?
- Is it a problem with their existing architecture than the architecture of the technology? In other words is it really bad design on their part and would it mean a lot of rework?

The thing to remember is that until recently, GUI developers were treated like second-class citizens. Real developers worked on the server and server work took priority. The designer needs to be sufficiently tech savvy to know when they are being deceived, and when they are not.

It's becoming easier for design groups to follow our lead, and hire designers with solid technical skills, and ideally also a background in HCI methods for user analysis and testing. Many kids design school students are already competent coders. Many engineering students are already aware of design and some are good designers. This kind of cross-disciplinary dexterity is exceptionally valuable in software design – think of its usefulness, for example, in discussions with engineers about creating web-based UI dashboards that are nimble, adaptive, and smartly designed.

Also invaluable: designers being willing to mentor developers who want to learn about design. At Citrix, our designers actively do this, inviting developers to join our integrated team on a part-time basis, helping to build components and product designs.

Through hiring, mentoring, and advocacy we are enabling the necessary and frankly inevitable rise of hybrid designers. The benefits are overwhelming: smarter and effective collaboration with engineering, more assurance of building improved products that match to the intended design specs, earning trust & respect with team partners that learn to adjust and improvise as issues arise, thus making the project much more satisfying and fulfilling when the release goals are achieved. And this approach boosts morale and overall improves product quality, thus elevating "design" in external team's eyes.

5 Conclusion

So are the inmates still running the asylum? At many companies, yes. In particular, engineers continue to dominate software product creation by controlling UI technology decisions, and, too often, by making these decisions for reasons that have little or nothing to do with the goal of giving users a great experience. As mentioned above, this is a pervasive problem in the software industry, thanks in part to the explosion in the number of available UI technologies.

At Citrix, we have dealt with this and related problems not simply by taking this decision out of the hands of engineers. Rather, we have taken steps systematically to broaden the role of designers in the product-creation process, while ensuring that they will collaborate with engineers and other stakeholders on a continual basis. We have also sought to ensure that this collaboration will be fruitful, by hiring designers who can code, and by actively seeking opportunities to work, formally and informally, with engineers who want to learn more about design.

We are very much still in the middle of this effort. But already, we see that the benefits of our approach have been noticeable and positive:

- Alignment of design and engineering activities with overall product development cycles
- General acceptance of a user-centered design process that has "teeth" when it comes to revisions or pushback by engineering

- Designers have an enhanced sense of influence, and a feeling of being more respected by their colleagues, particularly engineers

More generally, we have built a stronger, more collaborative relationship with Citrix's engineers. Both parties are now more trusting of one another, are better able to make adjustments on the fly, and feel they can achieve their own goals while helping others do likewise. The result is not just better morale, but also, critically, products that are better because they more closely match design specs, while maintaining the technical sophistication Citrix is known for. And "design" is now known, within Citrix, as much more than pretty window-dressing – as, indeed, a critical element of any effort to make great products. Obviously this has not happened overnight, and we still have much to do. But we are very hopeful that we will continue to see success in this area.

Action Items. For readers facing similar situations, what are some next steps you can take to start the process of balancing ownership of UI technology and forming equal respected partnerships with engineering?

- Start reading up and learning popular UI technologies and tools; make it a quarterly goal with 20 % of your time dedicated to this. (like Google's famous 20% time to do your own projects, which works out to be a single day per week)
- Learn the platform UI guidelines. It really helps when you need to back up what you are telling an engineering team when you can turn to published guidelines from Microsoft or Apple!
- Participate (or at least listen in) on engineering discussions, noting key concepts and phrases and issues. Tap your social network on people and resources that can help you advance your learning.
- Try out sample projects, taking your own designs and making them interactive with technologies like Flash or Silverlight or AIR. Guides, tutorials, and videos are online.
- Attend developer conferences like Microsoft MIX or Adobe MX which often hold hands-on training sessions with sample projects. Map those lessons to what you're doing at work, and take it further.

Reference

1. Cooper, A.: The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams Pearson Eduction (2004)

Connecting Usages with Usability Analysis through the User Experience Risk Assessment Model: A Case Study in the Tourism Domain

Alessandro Inversini[1], Lorenzo Cantoni[1], and Davide Bolchini[2]

[1] webatelier.net, School of Communication Science, Università della Svizzera Italiana, Lugano, Switzerland
[2] User Simulation & Experience Research Lab, School of Informatics Indiana University, USA
alessandro.inversini@usi.ch, lorenzo.cantoni@usi.ch, dbolchin@iupui.edu

Abstract. Web usability evaluation methods have been traditionally considered as detached from the analysis of the actual usages of a web applications. While the former is typically delegated to web engineers or web designers, the latter seems to be a concern only for online marketing experts. Based on our previous research results, in this paper we present a holistic evaluation model that seamlessly integrates usability and usage analysis in the assessment of the communication quality of a web application. Specifically, we apply this model to the analysis of BravoFly website (a Swiss Online Travel Agent) and we illustrate how the results of this integrated evaluation can shed new light in intelligently prioritizing re-design interventions. Implications for online tourism communication managers and researchers in this area are discussed.

Keywords: usability evaluation, usability inspection, usability testing, usage analysis, log files, design dimensions.

1 Introduction

Tourism is an information-intensive domain [1] which has been completely reshaped by the advent of the Internet [2]. In general, companies in the tourism business are using and exploiting the Internet with two main goals: (1) to market themselves online [3]; and (2) to sell tourism products through the Web [4]. At a closer look, the tourism online domain [5] is composed by different players (such as different online travel agents) trying to compete for the user attention [6] to market and sell tourism products (e.g. flights). In this process, the quality of online communication [7] [8] in terms of usability at each level (from content quality, navigation quality, transaction quality and overall usability) has become a fundamental issue in the field. To complement usability, it is also important to consider the role of usage analysis within the domain: different researches in the last years are trying to devise techniques to make sense out of the huge amount of usages data (e.g. coming from log files) generated by tourism websites [8]. In this respect, there is an increasing effort in focusing on the analysis of

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 283–293, 2011.

communication issues related to usages and online behavior [9]. Traditional usability and usage analysis, however, are not yet recognized as part of a meaningful and coherent theoretical framework. In our previous research [7] [8] we have laid the foundations for the User Experience Risk Assessment Model, an attempt to connect usability analysis with usage analysis through the unifying notion of risk. On the one hand, usability problems identified though usability methods can be considered significant risk factors for a detrimental user experience. On the other hand, results from usage analysis identify the probability for users to be actually exposed to those usability problems, thus mitigating or worsening the overall risk for negative experiences. Based on this theoretical elaboration, a proper analysis of the user experience risk would inform project managers, communication and web designers in making decisions concerning questions such as: what parts of the application require immediate attention for re-design or improvement? Are my users exposed to potentially negative experiences? How can I optimize the good experiences on my site? Our innovative contribution is the elaboration of few, basic constructs to analyze and characterize such hurdle of risk issues by holistically leveraging current approaches to usability analysis and usage studies. More analytically, our research proposes to see the user experience risk as composed of three main elements: (i) threats as usability problems inherent to the design; (ii) vulnerability as the exposure to usability problems and (iii) resilience as the user's ability to overcome usability problems.

A case study approach has been used to investigate and validate the model. We chose Bravofly.com as a representative case of information-intensive web application in the Tourism domain, and specifically on the online travel/flight business. BravoFly is a Swiss based company which mainly operates as Online Travel Agent in south Europe (7 countries + international version) for low cost flights. BravoFly is strongly devoted to innovation seeking competitive advantages and its unique aspects could be seen in the possibility of combining low cost carriers on different destinations to get price advantages for customers.

2 Related Works

According to Garrett [10], "user experience is not about how a product works on the inside (although that sometimes has a lot of influence), but it is about how it works on the outside, where a person comes into contact with it and has to work with it". The same author described the website as a "self-service product", where no instruction manual or seminar is provided: the user faces the website alone, only with her/his experience guiding her/him [10]. Furthermore, Kuniavsky [11] investigated the concept of user experience identifying three main factors that positively affect user experience, namely (i) functionality, which considers the websites' usefulness with regard to the users, (ii) efficiency, which considers the time needed by the users to accomplish specific tasks, and finally (iii) desirability, which considers the users' feelings of surprise and satisfaction with regards to the web application. ISO (International Organization for Standardization) defines usability (ISO 9241) as "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments". The various aspects of this definition are also supported by Cantoni and Tardini [12] which define usability according to the Website Communication Model (in short WCM) as "the adequacy of contents/functionalities (pillar I [of the Website Communication Model - WCM]) and accessibility tools (pillar II), between themselves and with respect to the users (pillar IV) and the relevant context (world). However, this adequacy has to be measured taking into consideration the goals of people who commission, project, develop, promote and run the website (pillar III)" (Cantoni and Tardini, 2006: 129-130). Besides, usage analysis (or log files analysis) is one of the most interesting studies to be performed on a website if there is no possibility of involving users during the usability analysis [13]. In general terms, log files are the traces left by the user while visiting the web site; this specific group of files are server side files that record users' activities while they are visiting the website. The study of the log files is not an engineering activity as such: log files analysis can give interesting information at a communicative level [9] such as the study of the users' paths along the website [14] by which it is possible to optimize the communication flow within the website. Tourism websites and moreover Online Travel Agent websites host and offer services or products (or in other words, gives visibility to third party websites). This websites should be well-designed and have great performances in order to satisfy both investors and the product/services providers and end-users. Moreover performance and conversions are critical success factors for eCommerce websites such as Online Travel agents websites. Good website usability normally leads to a good website performance; therefore usability performance is a key success factor for a website [15], [16].

2.1 Theoretical Framework

Understanding the user experience with Online Travel Agent websites is a daunting and multifaceted task. The theoretical and methodological foundations developed by the community of scholars and design professionals in many fields (human-computer interaction design, usability, interaction design, marketing, software engineering) to tackle various aspects of the user experience have remained very isolated and self-contained. The consequence of such conceptual fragmentation has been a proliferation of methods and techniques that lacks a comprehensive, holistic perspective on the issues at stake. In particular, there are two areas which have seldom dialogued with one another: the study of usability and the study of usages. On the one hand, usability studies have typically focused on the empirical evaluation of the efficiency and effectiveness of the website to support user goals and tasks, with the aim of improving the quality of the design [17,][18]. On the other hand, the study of usages has mainly addressed the analysis of website traffic, aggregated user's paths and ecological factors (referrals, in coming and out coming websites), with the purpose of informing marketing actions and visibility [13]. The common, unifying factor among these two areas of concerns is the study of the user as a person exposed to a complex, articulate – and oftentimes unpredictable – communication artifact. Interestingly, this type of situation is not at all exclusive or uniquely distinctive of electronic communication. The study of analyzing, evaluating and predicting the consequence of a person's exposition to potentially adverse events is common to many other disciplines, including – just to name a few – public health, security engineering, emergency management, and finance [19], [20]. In their theoretical frameworks, these disciplines have always leveraged a basic construct: the notion of risk. The relevant components, and formulae, defining risk vary from discipline to discipline, given the different type of problems to solve and analyze. We therefore do not review all the possible combinations and variants of conceptual constructs defining risk in the abovementioned disciplines. For the sake of our theoretical elaboration, however, we have identified some important, common factors determining risk which are readily applicable to the study of the user experience in interactive communication. The design and deployment of a destination website can be considered an enterprise which is subject to some degree of risk: actual users are often unknown (although possibly predicted during design), the actual behaviors of the users on the site is often unknown from the outset, and the actual effect or outcome of the experience with the site on the user is difficult to predict. Most importantly, the complexity of the design features of large web applications (and their emergent properties due to their interconnectedness) poses additional levels of unpredictability to such factors, augmenting the risk of negative user experiences. A proper analysis of the user experience risk would inform project managers, communication and web designers in making decisions concerning questions such as: what parts of the application require immediate attention for re-design or improvement? Are my users exposed to potentially negative experiences? How can I optimize the good experiences on my site? Our innovative contribution is the elaboration of few, basic constructs to analyze and characterize such hurdle of risk issues by holistically leveraging current approaches to usability analysis and usage studies. Overall, the risk of negative user experiences with the site is determined by 3 basic factors: threats, vulnerability and resilience (see Fig. 1), which are explained in detail in the following paragraphs.

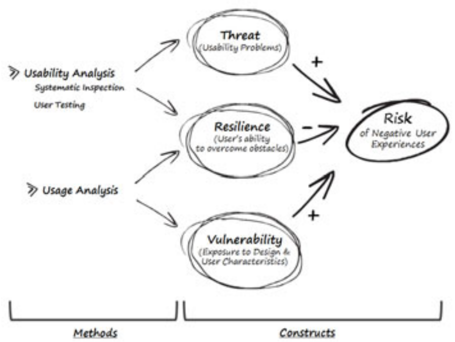


Fig. 1. The synergy of usability and usage analysis to capture user experience ris

"Threats" as Usability Problems Inherent to the Design. The design complexity of large destination websites is often prone to usability problems. Although there are many definitions of usability problems, for the purpose of our framework, we define a usability problem as a design defect that is a potential threat to an optimal user experience. Usability problems of varying severity are typically inherent to how the application has been designed and, therefore, eventually lie at one or more of the following design dimensions:

- Content: the core information messages of the websites, from text, to multimedia. An example of potential threat, or usability problem, at this level is the presence of non obsolete content, or the absence of contact information.

- Information architecture: the overall organization of the content in chunks and sections. An example of potential threat at this level is the classification of the content using a limited set of criteria (e.g. only by geographical location) which do not correspond to the user's natural reasoning in exploring information (e.g. I want to go skiing, no matter the specific location).
- Navigation and interaction: the strategies by which users can use and move around the information architecture through links and interact with content. An example of threat at this level is the lack of intuitive mechanisms to use an interactive map, to navigate to in-depth information from it and to print the desired information.
- Services & transactions: the strategies by which specific operations and services are organized, structured and made accomplishable by the user.
- Search functionality: the way an internal search engines supports accurate and efficient retrieval of information.

Labeling and interface semiotics: the way in which all the above mentioned aspects are conveyed at the interface level through naming conventions, layout strategies, metaphors and labels. This analytical modeling of the threats reveals critical areas of the site that do not necessarily determine a risk of negative experiences, but need to be carefully and jointly considered with respect to the vulnerability that our users may manifest.

Vulnerability" as Exposure to Usability Problems. An area of the site (a list of hotels) with severe usability problems (out-of-date or missing contact information) may not be considered a too dangerous threat if, for instance, no user ever bumped into it (because it was completely buried in the site architecture). The fact that the list of hotels is very difficult to find mitigates the potentially destructive effect of the threat because the actual exposure of our users to it can be considered very low or null. This simple example shows that user's vulnerability to a threat can be defined as the exposure of the users to it, identified in terms of actual traffic or potentially accessible pathways. Of course, the fact that few users access the list of hotels has to be carefully analyzed: is that in line with overall website goals or that should be the most important website area? Are there usability problems (most probably in the navigation layer) that prevent users from reaching it? Are promotional activities bringing the most appropriate publics to the website? In addition, vulnerability also depends on the specific characteristics of our users, which may be more or less sensitive to a threat. For example, web-savvy users may find no problem in downloading an additional Flash player to enjoy a video. Senior citizens new to web technology, on the contrary, may have a hard time in figuring out what to do in this situation. The consideration of these elements determines important factors that influence the degree of vulnerability of our users.

"Resilience" as the User's Ability to Overcome Usability Problems. Whereas vulnerability identifies the danger of a potential or actual exposure to a threat, risk can be highly mitigated by considering how and whether users actually overcome – or "survive" to (in virtual sense) – a threat. Let's assume that users often visit the section to subscribe to the newsletters, and they go through a cumbersome set of poorly organized pages to create an account, necessary to subscribe to the newsletter. The fact that 90% of the people accessing the newsletter section are eventually able to complete the task is a clear sign of high resilience, which is the capability of the users to overcome obstacles posed by existing threats. And this mitigates the overall risk of negative user experiences. As the user population changes, however, this resilience may radically and suddenly vary, causing a high level of risk.

The Synergy between Usability Analysis and Usage Studies. In this perspective, the existing methods and approaches to usability analysis and usage analysis can work in concert to address the key issues outlined by our user experience risk framework. Usability analysis (through systematic inspection and user testing) can unearth the threats for the user experience. The main outcome of usability analysis is, in fact, an organized set of usability problems inherent to the design. On the one hand, user testing can also uncover an important aspect of resilience (observing whether and how users are affected or overcome usability problems). On the other hand, structured inspection methods such as scenario-based inspection and cognitive walkthrough (common practices in interaction design and usability engineering) can reveal other interesting aspects: the characteristics of the user profiles identified and considered throughout the inspection process enable to reason about vulnerability to the usability problems identified. Usage analysis perfectly works as a complementary analytical toolkit to determine vulnerability in terms of intensity of traffic to poorly designed areas. Moreover, the study of usages can be applied to reveal episodes of resilience, as the analysis of the full paths of the user and their conversion rate is taken into consideration. The following sections articulate the application of our approach to the evaluation of the user experience risk analysis of an online travel agent website (i.e. www.bravofly.com). The purpose is to show how this conceptual framework can translate into an analytical instrumentation which can be organically used to inform new discoveries in the study of the user experience and to enable better designs.

3 Research Design

Bravofly.com and Online Travel Agency has been investigated over the course of 1 year by investigating its website using the following methodologies: heuristic driven evaluation, scenario based user testing (n=16) and usages analysis (1 year time frame). While (i) heuristic driven evaluation [16] is a methodology that evaluate possible usability drawbacks thanks to the efforts of one expert evaluator (also called usability inspector) and to a set of given heuristics (i.e. usability guidelines), (ii) user testing [19] is a methodology that involves possible real users confronted with the live application facing with real scenarios, goals and tasks to accomplish. These two analysis were based on MiLE+ (Milano – Lugano Usability Method – [21] [22] [23]). Finally, the (iii) usages analysis, which is not only an engineering activity but could give interesting insights also at communication level [9], have been studies thanks to the Google Analytics software installed by the company. Following MiLE+ methodology [21] an Usability Kit (Ukit in short) was elaborated in order to guide the heuristic evaluation and the user testing. The Ukit was composed by 5 user profile, 5 goals, 66 tasks and 33 heuristics divided into (i) content, (ii) navigation and (iii) graphic. The evaluator was asked to check the website against the heuristics; in order to make an ex-post reconciliation and to confront the results of the two analysis the results of the user testing were also associated where possible to the heuristics.

Finally, results (i.e. heuristic evaluation, user testing and analytics analysis) have been mapped on a light reverse design of the application. Thus the main goal that has been investigated within this research is the composition of user experience of three main elements: (i) threats as usability problems inherent to the design; (ii) vulnerability as the exposure to usability problems and (iii) resilience as the user's ability to overcome usability problems.

4 Results

Results have been divided following the main methodologies used within the research.

Expert Review: the expert review found 75 usability breakdowns grouped as follows: 28 issues referred to content aspects (number of heuristics 7), 17 referred to navigation aspects (number of heuristics 12), 30 referred to graphic aspects (number of heuristics 13). The three most recurrent usability breakdowns are:

- Content: Accuracy (frequency 14/75).
- Navigation: Accessibility (frequency 6/75).
- Graphic: Font (frequency 7/75).

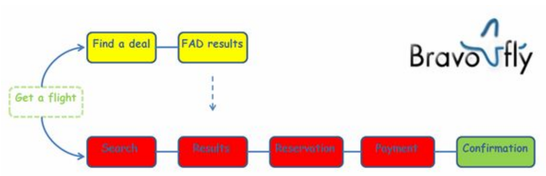


Fig. 2. Expert Review Reverse Design

This picture (Figure 2) shows the distribution of the 75 usability errors found with heuristic evaluation within the light reverse design of the website.

Sections errors frequency: Search (11), Results (19), Reservation (17), Payment (11) Confirmation (3) Find a Deal (10) Find a Deal Results(6).

User Testing: 16 users have been involved in the user testing. Demographically they were divided in to 10 males and 6 females, aged from 19 to 36 with 5 different nationalities (Italian n= 7, Swiss n=5, German n=2, British n=1 and Byelorussiann=1). These prospective users were tested against 3 scenarios (48 situations and 61 tasks) with the methodology of thinking aloud. The study found 54 usability breakdowns grouped as follows: 19 issues referred to content aspects, 21 referred to navigation aspects, 14 referred to graphic aspects. The three most recurrent usability breakdowns are:

- Content: Accuracy (frequency 11/54)
- Navigation: Accessibility (frequency 6/54)
- Graphic: Icon Consistency (frequency 4/54)

This picture (Figure 3) shows the distribution of the 54 usability errors found with user testing within the light reverse design of the website.

Sections errors frequency: Search (11), Results (12), Reservation (9), Payment (7) Confirmation (3) Find a Deal (4) Find a Deal Results (8).

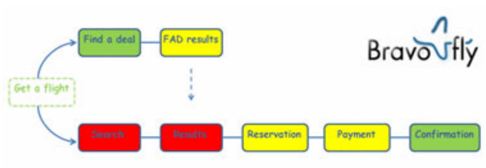


Fig. 3. User Testing Reverse Design

For what concerns the time in accomplishing the tasks (figure 4) it is possible to note that the task related to luggage rules information along with the ones about bravofly special combination (i.e. flight combinations only available on bravofly website) are very demanding for the end users. This is due mainly to the accuracy of the description within the sections.

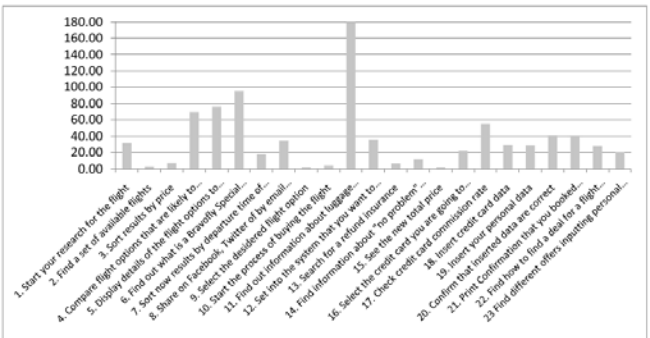


Fig. 4. User Testing task and Time

Log Files Analysis: the considered period is from December 14, 2009 - September 14, 2010. The website got 31,327,866 visits and 125,656,872 page views. Each user viewed 4.01 pages with a bounce rate of 24.34%; the average time on site was 00:05:14 per visit. Before going deeper in the log files evaluation had been necessary a process of data classification. In order to study the effective usages of the website, the research considered the measure of the website top contents as relevant. Thus, data coming from all the visits in all the sections were analyzed. Top contents (figure 5) are: flight results (45.8% of the total visits), home page (20.5% of the total visits) and passenger details (8.6% of the total visits).

Table 1. Confront usability issues and usages (NED= not enough data)

Sections	Heuristic Evaluation	User Testing	Usages
Search	11 issues	11 issues	20.5%
Results	19 issues	12 issues	45.8%
Reservation	17 issues	9 issues	NED
Payment	11 issues	7 issues	8.6%
Confirmation	3 issues	3 issues	NED
Find a Deal	10 issues	4 issues	NED
FAD Results	6 issues	8 issues	NED

Usability Risk Assessment. All the risk assessment detections have been compared and the conclusion is that crucial problems within the Bravofly's usability regard the "Results" section, followed by the "Search". This result totally reflects the log files analysis trends, and hence it is possible to assume that it is because those are the sections where the users spend at most the time in order to carefully decide if buy, or where and when to fly thanks to Bravofly.com. Combining the three analysis (table 1) it is possible to assess the risks connected with the user experience for the Bravofly website. Usability analysis highlight that form the expert point of view the threat are mostly in the results, reservation and search sections, while from users seems to have less problems within the reservation section, overcoming some of the problems highlighted in the heuristic evaluation. Finally, vulnerability as reflected by analysis of the usages is mostly in the results page and in the search/home page. Thus risks are high for these sections (i.e. results and search/home) as they are the ones in which there is an high concentration of threats (results expert evaluation), low resilience (results from user testing) and high vulnerability (results of the usages analysis).

This picture (figure 6) shows the distribution of the website visits within the light reverse design of the website.

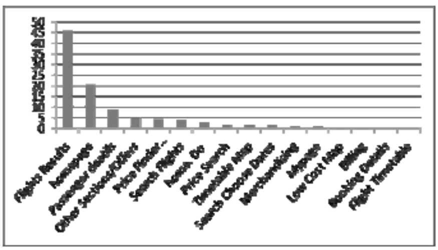


Fig. 5. Distribution of the visits to the website

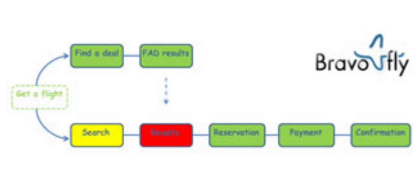


Fig. 6. Distribution of the visits to the website reverse design

5 Conclusions and Limitations

Results show how the use of the User Experience Risk Assessment Model and the systematic connection between usability and usages properly inform website redesign by well balancing business and user experience goals. Starting from this analysis, the implications for company managers are clear: it is possible to invest a fully application redesign informed by the usability analysis (both heuristic inspection and user testing), but it is also possible to concentrate efforts and resources within those sections which are the most visited. Limitations might be listed on two different areas: (i) technical: the usages analysis only considered the most visited pages due to the limitation of the tool used (i.e. Google Analytics) and (ii) structural: the model does not really considers the sales or management objectives. In order to overcome the first limitation log files raw data should be used in order to have a wide spectrum of actions to be performed on the usages (e.g. most frequent paths, advanced IP filtering etc.). In order to overcome the second limitation managers' semi structured interviews might be added to the model to better focus on sales objectives. Actually the percentage of visits of a given section could also be influenced by the promotional activities performed on the website.

References

[1] Poon, A.: Tourism, Technology and Competitive Strategies, Wallingford, CT. CAB International, Oxford (1993)

[2] Buhalis, D.: eTourism: Information technology for strategic tourism management. Prentice Hall, Harlow (2003)

[3] Buhalis, D., Spada, A.: Destination Management Systems: Criteria for success. Information Technology and Tourism 3(1), 41–58 (2000)

[4] Werthner, H., Klein, S.: Information Technology and Tourism – A Challenging Relationship. Springer, Wien (1999)

[5] Xiang, Z., Wöber, K., Fesenmaier, D.R.: The representation of the tourism domain in search engines. Journal of Travel Research 47, 137–150 (2008)

[6] Inversini, A., Buhalis, D.: Information Convergence in the Long Tail. The Case of Tourism Destination InformationIn. In: Hopken, W., Gretzel, U., Law, R. (eds.) Proceedings of the International Conference on Information and Communication Technologies in Tourism 2009, Amsterdam, Netherland, pp. 381–392. Springer, Wien (2009)

[7] Inversini, A., Cantoni, L.: Cultural Destination Usability: The Case of Visit Bath. In: Hopken, W., Gretzel, U., Law, R. (eds.) Proceedings of the International Conference on Information and Communication Technologies in Tourism 2009, Amsterdam, Netherland, pp. 319–331. Springer, Wien (2009)

[8] Inversini, A., Cantoni, L., Bolchini, D.: Presenting UsERA: User Experience Risk Assessment Model. In: Gretzel, U., Law, R., Fuchs, M. (eds.) Proceedings of the International Conference on Information and Communication Technologies in Tourism 2010, Lugano, Switzerland, pp. 99–110. Springer, Wien (2010)

[9] Cantoni, L., Ceriani, L.: Fare comunicazione online, analisi dell'attività di un sito internet attraverso i file di log. Comunicazione Italiana, Roma (2007)

[10] Garrett, J.J.: The elements of users experience. AIGAl, NY (2003)

[11] Kuniavsky, M.: Observing the User Experience: A Practitioner's Guide to User Research. Morgan Kaufmann, San Francisco (2003)

[12] Cantoni, L., Tardini, S.: Internet. Routledge, London (2006)

[13] Atterer, R., Wnuk, M., Schmidt, A.: Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In: Proceedings of the 15th International Conference on World Wide Web, Edinburgh, Scotland (2006)

[14] Pitkow, J.: In search of reliable usage data on the WWW. In: Sixth International World Wide Web Conference, Santa Clara, CA, pp. 451–463 (1997)

[15] Douglas, A., Mills, J.E.: Staying Afloat in the Tropics: Allying a Structural Equation Model Approach to Evaluating National Tourism Organization Websites in the Caribbean. In: Law, R., Mills, J.E. (eds.) Handbook of Consumer Behavior, Tourism and the Internet, pp. 269–293 (2004)

[16] Nielsen, J.: Usability inspection methods. In: Conference on Human Factors in Computing Systems, pp. 413–414 (1994)

[17] Au Yeung, T., Law, R.: Usability Evaluation of Hong Kong Hotel Websites. Information and Communication Technology in Tourism, eds. City (2003)

[18] Brinck, T., Gergle, D., Wood, S.D.: Usability for the web. Morgan Kaufmann, San Francisco (2002)

[19] Adams, J.: Risk, p. 228. University College London Press, London (1995)

[20] Blaikie, P., Cannon, T., Davis, I., Wisner, B.: At Risk: Natural Hazards, People's Vulnerability, and Disasters, pp. 333–352. Routledge, London (1994)

[21] Triacca, L., Bolchini, D., Botturi, L., Inversini, A.: MiLE: Systematic Usability Evaluation for E-learning Web Applications. In: Proceedings of EDMEDIA 2004, Lugano, Switzerland, pp. 4398–4405 (2004)
[22] Triacca, L., Inversini, A., Bolchini, D.: Evaluating Web Usability with MiLE+. In: Proc. Seventh International Symposium on Web Site Evolution (WSE 2005), Budapest, September 22-29 (2005)
[23] Bolchini, D., Garzotto, F.: Quality and Potential for Adoption of Web Usability Evaluation Methods: An Empirical Study on MILE+. Journal of Web Engineering 7(4), 299–317 (2008)

Ethnographic Research of User Behavior of Mobile Devices of China, Korea, India, and The Netherlands

Daeeop Kim and Kun-Pyo Lee

Industrial Design, KAIST, Yuseonggu, Daejeon, South Korea
{up4201,kplee}@kaist.ac.kr

Abstract. A product and its cultural understanding explain the social perception and value of the product and its users. Recent cultural studies show that cultural differences can have influence on the usage and satisfaction of the product through the advent of the Internet and globalization. This study explores such differences in four countries, based on observations on the usage of mobile media. As part of ethnographical method, researchers visited Korea, China, India and the Netherlands, observed and interviewed a total of 48 subjects (12 for each country) in order to analyze the characteristics their use of mobile media. The observations were performed over four sectors of gaining, managing, sharing and enjoying media content, which were further investigated through two times of comprehensive synthesized analysis by local researchers. After the observations, 10 major differences were found over three categories. How to collect and share media content varied depending on each culture. Particularly, content was used for personal entertainment as well as for social networking purpose with varied details. These differences are believed to stem from cultural differences, which would help understand the expected experience and value in each of the countries.

Keywords: User Centered Design, Cultural Difference, Interface Design, Ethnographical Research.

1 Introduction

1.1 Cultural Differences

The necessity of culture studies arises when human, as the agent of interface manipulation, and their cognition process are taken into consideration.[1] The main interest of HCI is to find the way to overcome the gap between the interface of the computing system and human being, thus, to offer a solution by which communication of manipulation and actual use is facilitated. The cognition and judgment of human is very important when it comes to interface manipulation. The user faces difficulty to manipulate a product if its interface doesn't meet the expectation of the users' mental model. Then, how is the users' mental model formed? According to Norman[2], the way the user experience the perception of the world affects the formation of this mental model. The users' experience may be

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 294–302, 2011.

regarded as the synthesis of memories in relation to everyday life, education, action, etc. and it cannot be separated from the cultural background and circumstances in which the user is living. Therefore, we think that the cultural background may also affect the user satisfaction upon the interface manipulation. Nisbetts's study proved several facts that can back up this kind of idea. He explained that the cultural difference between the East and West strongly affects the criteria of cognition and judgments; The way of people choose nouns, how they define the upper side and the lower side, and how they differ themselves and others. [3]

To figure out the cultural difference focusing on the utilization culture and the value sets, like the case of Boztepe, understanding upon Cross Cultural Communication is necessary.[4] We think it will be helpful for designing a product which, beyond providing fundamental usability, satisfies the overall value and requirements of the user. Furthermore, the recent trend of design objective is changing to 'providing a more amazing experience' from 'maximizing the efficiency of it.' Due to this kind of change, studies in relation to the user experience are becoming more important and interests toward user participating design method are arising since it may be useful to collect and analyze the user's experience.

1.2 Several Issues That Describe Cultural Differences

Researchers including Nisbett[3] suggested comparison key words that showed strong contrast among each culture. In case of Hofstede[5], he explained that there is an invisible part and clearly distinctive part between the individuals and culture, and in human nature. Above the common nature that triggers human behavior, culture is positioned and on the top of it the personal characteristics exists. Studies upon cultural difference have been strenuously focusing on the difference between these two areas. Accordingly, in this study, rather than focusing on the common behavior in the use of a given product, we focus on the difference between personality and culture.

Hofstede made use of culture dimension to define work-related value. The reason why we chose his argument was because it provides a viewpoint by which we can understand the way that the participants perform a given task that is mainly about the interaction with a product. We chose Hall's dimension to find out what releases the right response. Through his argument, we expected to decide the appropriateness of each behavior against using a product. We chose Trompenaars's dimension to determine the way in which a group of people solves problem.[6] Through his argument, we expected to gain understandings about the participants' objectives of using the product.

In short, to understand the cultural difference regarding the product, we covered the purpose of the use, the influence of the product use by each culture, and the cultural tolerance shown for the product. However, we didn't expect that all the comparison criteria would be helpful. As we didn't try to find the answers for many questions nor to observe various kinds of behaviors but only those that were limited and related to the use of product, the analysis were made selectively with the cultural dimension that we thought would be proper for this study.

Table 1. Synthesized categories of Cultural Variables from Hofstede, T.Hall and Trompenaars

Researchers	Cultural Variable Categories
Hofstede	Power Distance
	Individualism vs. Collectivism
	Masculinity vs. Feminity
	Uncertainty avoidance
Edward T.Hall	Time
	Context
	Proxemics
	PMS
	Message Velocity
	Action Chain
Trompenaars	Universalism vs. Particularism
	Individualism vs. Collectivism
	Neutral vs. Emotional
	Specific vs. Diffuse
	Achievement vs. Ascription
	Attitudes Toward Time
	Attitudes Toward Environment

2 Observation Research

Along with cultural gap, many other factors were considered in selecting four countries. These candidates were finally boiled down to Korea, China, Indian and the Netherlands by availability of researchers, subjects with similar standard of living, Internet accessibility, and the prospect of cultural research. Korea was an initiative choice, because it was where the research company was located, the research was designed, and allowed more pilot tests to be attempted. The major reason for China was that it is one of the fastest growing economies together with Brazil, Russia and India (so-called the BRIC) and has the biggest market in the world. India was chosen because it was also one of the BRIC members. Since Indians shows widely ranging gaps depending on the class, region, and religion, the Indian culture cannot be defined simply. However, India has shown a remarkable economic growth lately, and few studies have done on its younger generation compared to other countries. The Netherlands made the list because it could present a stark cultural contrast with those of Korea, China and India, helping the study to reveal their cultural environments. This study mainly used a process of observing users' behaviors and analyzing them. The observations first were performed as a primary approach, and in-depth interviews and brief survey were added to supplement the interpretation. In case a country-specific obstacle might arise, the field study was divided into required and optional sections to provide Procedural flexibly during the limited study period.

The results of the observations and interviews were categorized by keywords with the KJ method. As they were organized into similar issues by an inductive approach, the differences became distinctive.

This study was based on comments from the 48 interviewees, 12 from each of the countries. Conclusion was made from these remarks, and the local researches re-organized the data through two times of workshops in order to exclude extreme interpretations. Because this study was qualitative by nature, how many people agreed or who said what would not be meaningful. Followings include the data verified by the local researchers.

The observation data were classified into 10 categories, which were sub-divided into three sections: personal aspect, social aspect, and technical trend.(Table 2.)

The personal aspect showed individual taste of media and products with a focus on a product, media, and personal experience and their value. The social aspect included a user's community surrounding the product and media content. The technical trend covered differences by country in social phenomena caused by trends of technical development.

3 Findings

3.1 Physical Possession and Management of Media Content

The Korean and Chinese users obtained high-definition videos and music through peer to peer methods. One of the most frequently used ways was an instant messenger, which was particularly apparent among the users in their early twenties. Since the Internet infrastructure was well and widely established in Korea, many of them enjoyed a music streaming service. They did not take a physical player as a requirement.

On the contrary, most of the Dutch interviewees said that it was somewhat special to buy media files, but that they were willing to purchase content for their favorite singers or ones worth an eternal storage.

There were no distinctive cultural differences in managing media. It is believed that how they keep and manage media files depend on personal taste and the development of the Internet. (Table. 2)

3.2 Use of Digital Media Product and Social Relations

A noticeably different set of data on awareness of technology was collected from the Chinese and Dutch users. The Chinese users thought it was good to buy a latest product with advanced functions. They regarded a gadget quipped with a variety of features as a good machine. Additionally, buying a much-hyped product could represent their self-images to others. The Dutch consumers agreed that a versatile product is good, but they did not feel the necessity to keep up with technological advancement and trends. One of the Dutch users even replied that it was hard for him/her to understand why Koreans or Chinese are so interested in buying newly released mobile phones. And most of the users from any country, except for China, preferred an easy-to-use product. But, some of the Chinese subjects said that they would choose products with as many features as possible, even if they end up not using all of them.

The Chinese, Korean and Indians thought that playing music out loud was a rude behavior. However, listening music without an earphone was seen socially accepted in China. The low-income Indian users were generous toward the noise. The Indians valued personal indulgence, the Koreans and Chinese appreciated respect for a person and his/her communities, and the Dutch interviewees priotized the obedience of social standards. The Indians showed a different approach toward privacy depending on their income status. For most of the low-income users, a mobile phone did not belong to one person. They believed that a phone could be shared with friends or family members whenever a phone chip is replaced. However, the middle- and high- income Indians emphasized privacy.

Table 2. Synthesized issues after affinity diagramming (KJ Mapping) with findings of user researches

	Categories
Physical Possession and Management of Media Content	Taste of Content Selection
	Characteristics of Media Usage
	Frequency of Media Usage
	Recognition of Copyright
Use of Digital Media Product and Social Relations	Social Awareness of Technology and Attitudes toward Convergence
	Etiquette and Privacy
	Personal Originality
	Importance of Instant Communication
	Photos and Entertainment

4 Analysis

The quantitatively collected research data can be analyzed on the basis of the previously suggested issues.

1. The issue regarding the product usage focused on cultural dimension
2. The cultural difference issue based on the contents use behavior
3. Experience values by different countries found from the synthesis of the two above-mentioned issues

4.1 The Issue Regarding the Product Usage Focused on Cultural Dimension

In the early part of this paper, we applied cultural dimensions suggested by other researchers to divide each area and to make a brief summary. With Hofstede's variable, we tried to figure out how it affects the interaction between the product and the user. With Hall's dimension, we tried to learn how appropriate each behavior, related to use of the product, could be. With Trompenaars's dimension, we expected to figure out the purpose that the participants have when they use the product.

However, on the contrary to our naive expectations, it wasn't meaningful to categorize the findings of this study into each area. Most of the findings were able to be divided into other categories. The conclusion we reached during a workshop session, in which the data was interpreted, was that the findings were either unessential or over-valued. But, luckily, some of the interpretations were still valid.

4.1.1 Individualism vs. Collectivism

None will show the difference between the East and West better than the case of Individualism and Collectivism. Whenever issues related to how people use contents or what/how they prefer to use occurred, Individualism and Collectivism showed a distinct contrast. In this study, the relationship between a person and his/her community surrounding media provided an important aspect. The Koreans, Chinese and Indians thought that using and sharing media was not just for the users, but also for friends or members in their communities. And by sharing their personal experience, the users confirmed and strengthened their ties to communities. Conversely, the Dutch users drew a clear line between the person and the group. This demonstrates that there have been needs for the huge trend of online social networking.

4.1.2 Uncertainty Avoidance

Korea, along with Japan, showed very high degree in avoiding uncertainty. If necessary, the participants tried to share as much information as they could. Although Korean male participants introduced themselves as having introversive personality, they were very active in communications, for example, exchanging SMS, file sharing, etc. At the time Hofstede ran his study, there would not have been enough data regarding China. However, based on the findings of this study, we assume that China will be in the similar range as of Korea for the case of this variable. In case of the Netherlands, the frequency of phone conversation or SMS was low, but, on the contrary, a lot of conversation and information was made and share during face-to-face communication.

4.1.3 Polychronic Time and Monochronic Time

Interestingly, the participants of this study showed quite a different results comparing to the existing studies. According to Hall, the Japanese participants appeared to be polychronic in personal relationship while to be Monochronic in work related situations. But the findings from our investigation showed that all the participants from Korea, China, and India preferred multi-functions both in relationships and performance; and due to their polychromic characteristics, preference toward multi-tasking performance were observed. It may be interpreted as the outcome of varied relationships and working conditions that have been witnessed along with technology development and diversity encouragement.

4.1.4 Low Context Culture – High Context Culture

In the aspect of communication, utilization degree of media, and sharing of media contents, the culture of each country showed significant differences in terms of the dimension of low context culture and high context culture. The Chinese and Korean participants were able to operate smooth communication with other groups even when the subject of conversation and behaviors were not clear. But the group with which they actively shared information was the more intimate group. Meantime, the Dutch participants were categorized to people from relatively low context culture. They preferred using direct and clear expressions. The Indian participants showed similarity in this part and it was more distinct in the higher income group.

4.1.5 Neutral vs. Emotional

In case of selection and utilization of media, the participants showed various layers of emotions which were quite different by each country. The Dutch participants reacted calmly and were rational upon their selection and utilization of media. But the Chinese participants identified themselves with the equipment and media they possessed. Accordingly, some of them revealed that they felt ashamed when their equipment reported error or malfunction.

4.1.6 Achievement vs. Ascription

It is well known that the people from the Eastern culture are ascription-oriented. While the Dutch participants did not identify themselves with the product, ascription of meaning was significantly observed among the participants from China, Korea, and India. They regarded the equipment that they paid a lot of money or a famous brand product as something that represented themselves. Some of the Korean participants purchased the equipment, as if it was a toy, and they answered they had bought it because of the certain functions it contains. They thought the equipment will represent their unique personality.

4.2 The Cultural Difference Issue Based on the Contents Use Behavior

This study found that despite the developmental gaps of a product and its infrastructure, there were cultural differences from country to country.

Those cultural differences in content can be summarized as followed. In terms of Obtaining content, all of the Koreans, Chinese, and Indians were open to purchasing favorite content and sharing it. However, the Korean and Chinese particularly enjoyed recording and creating their personal experiences, which was contrary to the Dutch users.

In regard to content Managing, personal taste and infrastructure, as opposed to cultural difference, were more influential. However, retouching photos and inventing unique poses were commonly found among the Korean users.

Regarding content Sharing, the Korean, Chinese and Indian participants used content sharing as a way to enhance their tie to their communities, or identify their positions in them.

In the cultural difference in Enjoying content, all of the users from the four countries put priority to personal satisfaction. However, how to utilize entertainment content as one of the social networking activities varied. In Korea and China, the content was used to represent their character, and in the Indian users emphasized entertainment as a source of amusement for multiple users. For the Dutch participants, entertainment content did not have any purpose other than personal satisfaction.

4.3 Experience Values by Different Countries Found from the Synthesis of the Two Above-Mentioned Issues

The type of experience the users expect with their mobile phones was also different. The Korean users thought recording personal experience and sharing it with others was the best use of mobile media. The Chinese emphasized personal satisfaction and social recognition, and the Indians priotized technical benefits a user and his/her group could enjoy, and the Dutch interviewees preferred a phone identifiable with them, based on privacy.

These differences revealed what users want with media products, and provided insight on how to position mobile products in local markets.

A mobile device that would be appropriate for the Chinese market is the one that has an appealing and unique identity. It seems like the Chinese users will prefer a mobile device which has multi-functions.

For Korean users, it seems like a device with which they can create and share experience will be the most appealing one. They prefer easy, useful, transformable, sharable, enjoyable product.

For Indian users, a product that can leverage a variety of contents will be useful. Relatively affordable price and a service system through which they can consume a lot of contents would be welcomed in their market. Also, providing a solution to allow a group of people share one device would be a realistic suggestion as well.

For Dutch users, it seems like practical and value-oriented product will be the appealing one. Throughout our investigation, we learned that they prefer to purchase products that would sustainably pursue identities related to socio-cultural issues (e.g., etiquette) or eco-friendly campaigns.

5 Conclusion

This study aimed to find cultural differences through the observation of the users' mobile equipment utilization. By applying the cultural dimensions, we found that the preferred values of each culture are different. The difference in value that we learned from this study may contribute to product design activities that are based on the users' experience. The findings will be a valuable data by which you can predict what kind of experience will be expected from a newly designed product in each culture. Moreover, it will provide useful information for localization strategy.

This study also showed the base of a useful methodology for culture-based research. It would be helpful if study designs were built from the cultural gaps that have been clarified through this Study. Public observation In the Netherlands and survey in India were particularly challenging. The Koreans were not active during interviews. These differences also have its roots in cultural attributes and need to be supplemented.

Acknowledgements. This research was designed and conducted with UDS Team, Jin-soo Ha and Jongwhan Seo. In China, we got help from Human Factors, Inc. In the Netherlands, Peter Jan Stappers and the Department of Industrial Design at TU Delft cooperated for local participants and their results analysis. And in India, it was

Raghu Kolly whom we should not forget to express our appreciation for cooperation of his team.

Also this research was supported by WCU(World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R33-2008-000-10033-0).

References

1. Alan Dix, J.E.F., Abowd, G.D., Beale, R.: Human-Computer Interaction, 3rd edn., Dix, A., (ed.) p. 834. Prentice Hall, Englewood Cliffs (2003)
2. Norman, D.A.: The design of everyday things, 1st Doubleday/Currency edn., xv, 257 p. Doubleday, New York (1990)
3. Nisbett, R.: The Geography of thought; How Asians and Westerners Think Differently and Why. The Free Press, New York (2003)
4. Boztepe, S.: Toward a framework of product development for global markets: a user-value-based approach. Design Studies 28(5), 21 (2007)
5. Hofstede, G.: Culture and Qrganizations: Software in the Mind. McGraw-Hill International, UK (1991)
6. Trompenaars, F.: Riding the waves of culture. McGraw-Hill, New York (1997)

A Conjoint Analysis of Attributes Affecting the Likelihood of Technology Use

Anna Elisabeth Pohlmeyer[1] and Lucïenne Blessing[2]

[1] University of Technology Berlin, Center of Human-Machine Systems, Franklinstr. 28/29, FR 2-6, 10587 Berlin, Germany
[2] University of Luxembourg, Engineering Design and Methodology, Campus Limpertsberg, 162A, avenue de la Faïencerie, L-1511 Luxembourg
anna.pohlmeyer@zmms.tu-berlin.de, lucienne.blessing@uni.lu

Abstract. Products are a composition of multiple attributes and should be evaluated in full-profile also in research contexts. Research studies on the use of technological products which only assess the importance of individual attributes do not reflect real-life scenarios of multi-attribute judgments and miss out on information that only becomes apparent in relative terms. In order to study the predictive value and relative importance of six attributes (functionality, cognitive ergonomics, classical ergonomics, quality, aesthetics, and emotional involvement) with respect to the likelihood of use, the method of conjoint analysis was borrowed from consumer research. The study was conducted with 104 participants divided in two groups of low and high self-competence. Group differences were only revealed when attributes were considered jointly, but not in single ratings. An intuitive interface, easy handling, and emotional involvement were greater motivators for the low competence group. Methodological implications are discussed.

Keywords: Technology Adoption, Self-Competence, Conjoint Analysis, Multi-Attribute Rating, Preferences.

1 Introduction

The consideration of what users want is central in user-centered design approaches. Also in marketing and product development, "the voice of the customer" has become an integral part of design specifications [1]. But what should be done if users want *everything*? If users are presented with a list of attributes such as good ergonomics or durable quality and asked how important these are, they might rate them all as very important. Without the necessity of trade-offs, independent ratings might lead to ceiling effects. But in everyday life, and also in product development as a special case, trade-offs are indispensable. Moreover, with too little variance in the responses, there is not enough information to differentiate preferences and requirements for different user groups. For example, both novices and experts might agree that ease of use is very important, but might prioritize this differently in relation to other attributes such as quality. Clearly, it is possible that different groups do agree on how important different attributes are. However, multi-attribute ratings are still necessary to validate

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 303–312, 2011.

the finding. The additional value and source of variance that can be derived from multi-attribute ratings is of interest here. The methodological challenge addressed is whether we gain more insight by rating products as a whole or by collecting ratings of their parts?

Concerning the definition of attributes, we follow Grunert [2, p.229]: *"An attribute can be defined as any aspect of the product itself or its use that can be used to compare product alternatives."* The attributes we address were identified in a previous study in which participants documented real-life examples of interactive technology over the course of one week and elaborated why they liked or disliked the products. In addition, product-independent reasons were given about what motivates the use of technology and what hinders it, respectively. These statements were categorized according to qualitative content analysis procedures [3]. The set of attributes included in the present study includes: *functionality, ease of use, ergonomics, quality, aesthetics, and emotional involvement* (see Table 1). Note that "*ease of use*" refers to the cognitive, information-processing side of ergonomics, while "*ergonomics*" is explicitly related to classical, physical aspects of ergonomics such as size and weight [4].

One prominent question in HCI touches the field of technology adoption: how should a system be designed to increase the likelihood of usage? In the present study we focus on usage scenarios but not on purchase situations. A purchase context with a monetary investment depends on consumer resources [5] and influences decision-making processes in a way that is not of relevance here. For this reason, 'financial issues' were excluded from the original list of attributes. Also, we tried to control for 'usefulness' by concentrating on one product, a digital camera, and on one scenario, namely hobby photography. The set of attributes was used to describe different models of digital cameras by varying the attributes within two levels (see Table 1). Participants were asked how likely they would be to use each camera for hobby photography. As in real life, all attributes were *considered jointly*.

Relative weights can serve as useful indicators for differentiating user preferences and thus setting priorities in early phases of product development as recommended by Ulrich and Eppinger [6]. In these early phases, user preferences have to be anticipated and are not derived from evaluating a final prototype. For this reason, established questionnaires such as Hassenzahl's AttrakDiff [7] that evaluate existing products and prototypes are no option for weighting user preferences at this stage.

Conjoint analysis is generally used on a very detailed level of product features. Here, a more abstract level of product attributes has been applied that can be transferred to other systems in future work. Although conjoint analysis is commonly applied in consumer research, purchase behavior was not of interest. Instead, the method was modified to assess anticipated technology adoption by asking for the likelihood of usage rather than acceptable pricing or purchase intent.

The most prominent model of technology acceptance was introduced by Davis in 1989 [8]. In essence, the technology acceptance model (TAM) claims that system use is best predicted by the intention to use the system. This, in turn, is influenced by the perceived ease of use and perceived usefulness of the system. The model has gone through numerous iterative modifications and extensions and dominates the field. Unfortunately, its wide acceptance does not equate to an equal amount of knowledge accumulation [9]. The approach in this study will not propose yet another TAM-modification. Instead, a methodological approach will be introduced to the field of technology acceptance. The method is not new by any means, but, to our knowledge, has not been applied at such an abstraction level and not in the context of technology adoption (with the criterion "likelihood of use"). In general, conjoint analyses are rarely seen in HCI despite their evident potential. In traditional TAM-studies only one system is being evaluated. Conducting a conjoint analysis allows the consideration of multiple system alternatives in the assessment. The resulting utility scores can predict preferences even of systems that have not been shown to the participants.

Finally, group differences with respect to self-competence (closely related to self-efficacy) with technology will be analyzed. Self-efficacy refers to an individual's belief whether they are able or competent to perform a specific task. Computer self-efficacy (ability to competently use computers) beliefs have been shown to be a valuable antecedent of technology use [10]. On the other hand, computer self-efficacy is also linked to consequences such as the affect (or liking) concerning computer use. An objective of the present study is to compare groups with self-perceived high or low technology competence with respect to relative attribute importance regarding technology use. We expect that the low competence group emphasizes attributes concerning the understanding and handling of technology (ease of use and ergonomics) more than the high competence group. It is also of interest how the outlook of positive affect (emotional involvement/ joy of use) will influence the likelihood of usage in the two groups.

2 Methods

The method of choice was conjoint analysis, and a full-profile rating in particular. A full-profile is a description of a product alternative including all attributes of investigation. Based on this information, participants rate several product alternatives. From the overall judgments, part-worth utilities are derived through regressional modeling. Thus, instead of computing a composite score of *single* attribute ratings, the path is a de-compositional calculation of attribute weights. Attention is focused on the relevance of attributes, rather than on differences between product models themselves [11]. The impact that a variation of attribute levels has on the likelihood of use will be expressed in relative importance values.

2.1 Participants

The sample consisted of 104 participants all living in Berlin, Germany. Two age groups were recruited: 52 younger adults (20-30 years, M_{young}= 25.88, SD_{young}= 2.73) and 52 older adults (65-75 years, M_{old}= 67.9, SD_{old}= 2.38). Each age group had equal numbers of men and women. The sample was well-educated; 43.3% had a university degree, and 26% had an "Abitur" secondary school qualification as their highest qualification. Younger adults were recruited through an online database of study volunteers. Older adults were additionally recruited through an advertisement in a weekly newspaper. Only volunteers who had used a digital camera previously qualified as study participants.

The sample was split into two groups regarding perceived subjective competence. Subjective competence is a five-item subscale of a German 19-item-questionnaire of technology affinity [12]. A median split conducted over the entire sample was confounded by age (more older adults felt less competent, χ^2 (1) = 9.85, $p < .05$; and Spearman's r (104) = -.3, p= .002). Therefore, median splits were conducted within each age group. As a result, young and older adults were divided equally between the groups. The two groups did not differ with respect to age (t (101.5) = .04, p> .05). However, as intended, the two groups differed regarding perceived subjective competence (t (94.5) = -11.48, p < .001). Also, the older subsample of the high competent group had significantly higher competence values than the young subsample in the low competent group (M_{young_low}= 3.3, SD_{young_low}= .56; M_{old_high}= 3.93, SD_{old_high}= .42; t (50) = -4.61, p < .001). This serves as proof that the grouping is valid and no longer biased by age effects.

2.2 Material and Task

Verbal product descriptions of different digital camera models served as study material. Models differed on six two-level attributes (see Table 1): *functionality, ease of use, ergonomics, quality, aesthetics, and emotional involvement*. The authors tried to define the levels in a way that firstly, made ranges between level 1 and 2 comparable across factors, and secondly, excluded knock-out criteria. Participants received thorough instructions and examples regarding attribute levels, which are not apparent in Table 1. For more elaborate description of the attribute levels see [13].

Table 1. Attributes and levels

attribute	level 1	level 2
functionality	primary functions	secondary functions
ease of use	takes getting used to	intuitive to use
ergonomics	handling requires physical effort	easy handling
quality	prone to defects; poor performance	reliable + durable; excellent performance
aesthetics	average appearance	appealing appearance
emotional involvement	not engaging, just functional	pleasurably engaging in addition to functional

Six attributes, each at two levels, allow 64 possible model combinations (2^6). Presenting all possible profiles would have been very tiring for the participants and consequently also endangering the results' reliability. Therefore, a model selection was made in order to reduce the number of models presented to a manageable amount. An orthogonal fractional factorial design was created using SPSS ORTHOPLAN. Orthogonal designs try to represent the entire set of models in a best possible way. The final design consisted of 20 model combinations (or so-called *stimulus cards*). A stimulus card is a list of the six attributes with a unique level-combination, describing a specific 'camera profile' (see Figure 1). There was no definite 'good' or 'bad' product description: each description was good with respect to some, but inferior with respect to other attributes.

Participants were asked to indicate how likely it was (on an 11-point-Likert scale ranging from 0% to 100% in steps of 10) that they would use the described camera for the purpose of hobby photography (see Figure 1).

Additionally, participants were asked to rate the importance of each attribute individually on a 10-point Likert scale (1 = unimportant; 10 = very important). This is what is referred to as *single* attribute ratings. A detailed, written elaboration of the attributes and their levels was provided and could be checked whenever necessary.

Finally, the 10-item version of the Achievement Motives Scale was assessed. An inventory that captures the two factors *hope of success* and *fear of failure* (5 items each) with 4-point Likert scales [14].

Fig. 1. Stimulus material and conjoint rating sheet

2.3 Setup and Procedure

In single sessions, each participant was instructed individually, following a standardized protocol. An entire session, including questionnaires regarding demographic information, participants' background in technology use and computer literacy, as well as a Kano analysis of user satisfaction, lasted 60-90 minutes. Participants were reimbursed with €10/hour. The data was collected using paper-and-pencil questionnaires.

The order of the product descriptions (stimulus cards) was randomized by shuffling the cards before rating. Prior to the rating of the camera models, participants completed the questionnaire on technology affinity and were asked for a baseline measure of how likely they would use a digital camera for hobby photography in general (on the same 11-point Likert scale of usage likelihood as provided for the product ratings).

2.4 Data Analysis

Metric conjoint analysis was employed to evaluate so-called part-worth utilities (analogous to weight estimates) and relative importance scores. Competence group differences were modeled as interactions within the regression. This was realized by dummy coding the groups and attributes as demonstrated by Bloch et al [15]. With this approach, between-group comparisons of regression coefficients can be tested statistically. The relative importance of each attribute is computed by calculating relative ranges (utility range of the attribute divided by the sum of the part-worth utility ranges of all six attributes). These values are percentages and consequently sum to 100. Data was analyzed using SPSS 19 software. α-level was considered at .05.

3 Results

The overall regression model was significant ($F\,(13, 26) = 224.43$, $p < .001$) and explained 99.1% of variance.

3.1 Predictive Value of Attributes

All attributes except functionality were significant predictors of usage likelihood. As expected, the prospect of intuitive use ($\beta = .58$, $t = 22.07$, $p < .001$), easy handling ($\beta = .70$, $t = 6.94$, $p < .001$), high quality ($\beta = .56$, $t = 21.36$, $p < .001$), an appealing appearance ($\beta = .18$, $t = 7.0$, $p < .001$), and emotional involvement ($\beta = .17$, $t = 6.33$, $p < .001$) increased likelihood of usage. The non-significance of functionality can be explained by equal numbers of participants who preferred the camera to have only primary functions integrated, who preferred the addition of secondary functions, and who were indifferent about this matter. This was revealed by a Kano analysis (see [13]).

3.2 Self-perceived Competence Group Differences

The groups did not differ with respect to baseline likelihood of using a digital camera for hobby photography ($M_{low} = 51.15$, $SD_{low} = 4.31$; $M_{high} = 59.23$, $SD_{high} = 3.84$; $t\,(102) = -1.40$, $p > .05$), which allows the straightforward interpretation of competence group differences.

There were no significant competence group main effects regarding *single* attribute ratings ($F\,(1, 102) = .23$, $p > .05$), nor group x attribute interaction effects ($F\,(3.6, 364.4) = .64$, $p > .05$; degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity). In other words, evaluating each attribute individually led to comparable ratings of the high and low competence groups (see Figure 2 and Table 2).

Part-Worth Utilities. Significant interactions of attributes and competence grouping could be observed with respect to ease of use ($\beta = -.23$, $t = -7.34$, $p < .001$), ergonomics ($\beta = -.11$, $t = -3.58$, $p < .05$), and emotional involvement ($\beta = -.12$, $t = -3.65$, $p < .05$). The low competence group demonstrated higher part-worth utilities of these three attributes (see Table 2).

Relative Importance Rating. The range of the part-worth utility values for each attribute is divided by the sum of utility ranges of all attributes. This provides a measure of how important the attribute is to overall preference. The size of the *relative importance* for each attribute indicates the impact that a level variation of this attribute has on the likelihood of use.

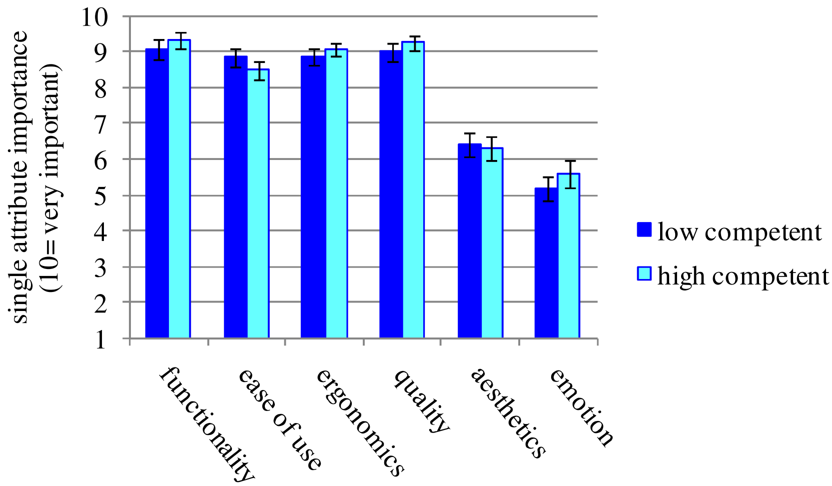


Fig. 2. Mean single attribute ratings (± *SE* mean) for low and high self-competence groups

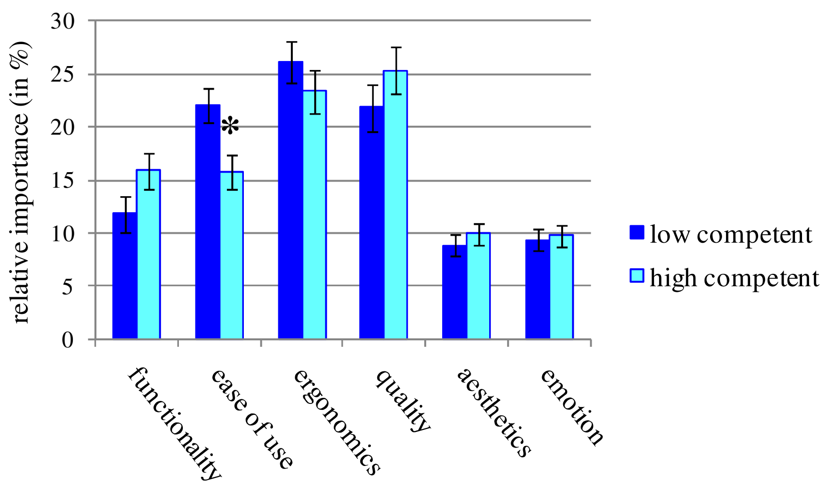


Fig. 3. Mean relative importance (± *SE* mean) for low and high self-competence groups

For the low competence group, (physical) ergonomics was the most important attribute while the high competence group's rating was mostly influenced by the degree of quality (see Figure 3 and Table 2).

Significant group differences were found for ease of use with a higher importance value for the low competence group (t (102) = 2.73, p< .05). However, no differences were found with respect to ergonomics (t (102) = .99, p> .05) nor emotional involvement (t (102) = -.30, p> .05)

Relative importance values can be interpreted in direct comparison because they are ratio scaled. For example, the variation of an intuitive usable interface in contrast to one that necessitates learning (ease of use) is twice as important as the variation of functionality (whether a device has only primary or additionally secondary features) for the low competence group (22.0 : 11.8). In contrast, the high competence group regards both attributes to be equally important (15.7 : 15.9).

Table 2. Overview of Single Importance Ratings, Part-Worth Utilities, and Relative Importance Values. Filled boxes indicate significant group differences.

attribute	group	single rating	part-worth utility	relative importance
functionality	low high	8.56 8.81	.927 .865	11.835 15.862
ease of use	low high	8.35 7.98	16.888 8.942	22.005 15.738
ergonomics	low high	8.35 8.56	20.619 16.750	26.104 23.329
quality	low high	8.50 8.75	16.350 17.442	21.831 25.300
aesthetics	low high	5.92 5.81	5.350 3.327	8.856 9.961
emotional involvement	low high	4.69 5.10	4.850 .904	9.368 9.809

Achievement Motives Scale. Participants in the low competence group showed significantly lower *hope of success* (M_{low}= 3.15, SD_{low}= .48; M_{high}= 3.33, SD_{high}= .56; t (101) = -2.13, $p < .05$) and a significantly *higher fear of failure* (M_{low}= 2.42, SD_{low}= .58; M_{high}= 2.12, SD_{high}= .60; t (101) = 2.62, $p < .05$).

4 Discussion

The likelihood of technology use, in this case a digital camera, depends on the combination of the product's attributes. There was consensus in the sample that an intuitive interface (cognitive ergonomics), easy handling of the device (classical ergonomics), high quality, an appealing appearance, and a pleasurable engagement increase the likelihood of usage. This list of attributes extends the two factors in the original TAM model [8]. Surprisingly, functionality did not contribute significantly to the prediction of technology adoption despite high *single* attribute importance scores. This was due to the operalization of functionality: the two levels (primary vs. secondary functions) were equally attractive.

Nonetheless, the model explained a near to perfect 99.1% of the variance. However, this should not be over-interpreted as the objective of this study was not to test a model, but to *compare* relevant attributes. Conjoint analysis is a method that decomposes overall ratings, in contrast to frequently used techniques of generating a composite score (e.g. sum) from independent ratings. The relative weights and importance values are of interest and not the overall rating itself.

As expected, the low competence group put more emphasis than their high competence counterparts on aspects of cognitive and classical ergonomics as these directly affect the probability of success or failure of an interaction. In addition, the prospect of an enjoyable interaction also increased the likelihood of usage to a higher extent in the low competence group. This should be considered as a promising motivator when designing for a hesitant user group such as a group with low self efficacy beliefs. As seen in the Achievement Motives scores, the low competence group is more likely to avoid anything that might lead to failures and has a lower tendency to aim for situations that might lead to success. Thus, any attribute that can encourage this group should be taken seriously. Emotional involvement is a hedonic attribute that is not directly linked to product performance in a pragmatic sense [16], but apparently to user motivation which is a prerequisite of interaction. In contrast, to get the high competence group interested, primarily high quality standards need to be met.

Group differences were not found in *single* attribute ratings, supporting our notion that additional information is gained when presenting products in full-profile rather than by independent attribute ratings.

One limitation of the study that should be mentioned is that the dependent variable was only a theoretically stated likelihood of usage. The link between such a predicted usage probability and actual behavior remains to be verified. On the other hand, this criterion creates the possibility to prospectively test expectations of user groups at very early stages of product development.

Conjoint analysis has been introduced as an effective approach to address the intention of technology use in HCI by considering attributes jointly. Attribute utility values from full-profile ratings and relative importance values showed group differences while independent importance ratings were not able to differentiate between users with low and high technology self-competence, respectively. Researchers should be cautious when relying on independent attribute importance ratings as these might yield to ceiling effects. An additional consideration of relative importance values is recommended. The derived preferences for different user groups can be translated into actionable priorities in a design process.

Acknowledgements. This research was supported by the German Research Foundation DFG [no. 1013/2 Prospective Design of Human-Technology Interaction] and by the National Research Fund Luxembourg [AFR].

References

1. Bradley, N.: Marketing Research. Tools and Techniques. Oxford University Press, Oxford (2007)
2. Grunert, K.G.: Attributes, Attribute values and their characteristics: a unifying approach and an example involving a complex household investment. Journal of Economic Psychology 10, 229–251 (1989)
3. Pohlmeyer, A.E., Blessing, L., Wandke, H., Maue, J.: The Value of Answers Without Question[s]: A Qualitative Approach to User Experience and Aging. In: Kurosu, M. (ed.) HCII 2009. LNCS, vol. 5619, pp. 894–903. Springer, Heidelberg (2009)

4. Hollnagel, E.: Cognitive Ergonomics: It's All in the Mind. Ergonomics 40(10), 1170–1182 (1997)
5. Blackwell, R.D., Miniard, P.W., Engel, J.F.: Consumer Behavior. Thomson/South-Western, Mason (2006)
6. Ulrich, K.T., Eppinger, S.D.: Product Design and Development. McGraw-Hill, New York (2007)
7. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer 2003. Interaktion in Bewegung, pp. 187–196. Teubner, Stuttgart (2003)
8. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–339 (1989)
9. Benbasat, I., Barki, H.: Quo vadis, TAM? Journal of the Association for Information Systems 8(4), 212–218 (2007)
10. Compeau, D.R., Higgins, C.A.: Computer Self-Efficacy: Development of a Measure and Initial Test. MIS Quarterly 19(2), 189–211 (1995)
11. Huber, J.: What We Have Learned from 20 Years of Conjoint Research: When to Use Self-Explicated, Graded Pairs, Full Profiles or Choice Experiments. Sawtooth Software, Research Paper Series (1997)
12. Karrer, K., Glaser, C., Clemens, C., Bruder, C.: Technikaffinität erfassen - der Fragebogen TA-EG. In: Lichtenstein, A., Stößel, C., Clemens, C. (eds.) Der Mensch als Mittelpunkt technischer Systeme. Berliner Werkstatt Mensch-Maschine-Systeme, vol. 8, pp. 196–201. VDI Verlag GmbH, Düsseldor (2009)
13. Pohlmeyer, A.E., Machens, F., Blessing, L.: Attractive or Not – What's the Difference? Inter- and Intra-Group Comparisons in the Kano Model. In: Marjanovic, D., Storga, M., Pavkovic, N., Bojcetic, N. (eds.) 11th International Design Conference - DESIGN 2010, pp. 413–422. University of Zagreb, Croatia (2010)
14. Lang, J.W.B., Fries, S.: A Revised 10-Item Version of the Achievement Motives Scale - Psychometric Properties in German-Speaking Samples. European Journal of Psychological Assessment 22(3), 216–224 (2006)
15. Bloch, P.H., Brunel, F.F., Arnold, T.J.: Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. Journal of Consumer Research 29, 551–565 (2003)
16. Hassenzahl, M.: The thing and I: understanding the relationship between user and product. In: Blythe, M., Overbeeke, C., Monk, A.F., Wright, P.C. (eds.) Funology: From Usability to Enjoyment, pp. 31–42. Kluwer Academic Publishers, Dodrecht (2004)

Personas on the Move: Making Personas for Today's Mobile Workforce

Michele Snyder, Anthony Sampanes,
Brent-Kaan White, and Lynn Rampoldi-Hnilo

Oracle Mobile Applications User Experience, 300 Oracle Parkway Office 326
Redwood Shores, CA 94065
michele.snyder@oracle.com,
{brent.white,lynn.rampoldi-hnilo}@oracle.com,.
csampane@cisco.com

Abstract. Personas are fictitious characters that are created to document the goals, behaviors, desires, and limitations of a real group of users. They are an important part of product development because they help keep focus on key user needs. Currently, most personas focus exclusively on desktop users from one geographic area. A new approach was used to create five personas that emphasize mobile usage and unique cultural traits. The personas were created after conducting ethnographic research in India, Singapore and the United States. The layout of the persona highlights the mobile work style, tasks, and use cases. Instead of setting all personas in a single country, each persona is localized and framed in a detailed cultural context. In addition, we included forward thinking aspects into our personas that projected what mobile users will be doing in the future based on expressed needs and the latest technologies.

Keywords: Persona, Ethnography, International research, Mobile.

1 Introduction

Personas have been used for more than a decade to influence the design direction of desktop applications. The concept of using a persona in the HCI field to aid in product design and development was introduced by Cooper [1]. Personas are fictitious characters that are created to document the goals, behaviors, desires, and limitations of a real group of users. The typical persona is a 1-2 page document that includes a name and picture, demographics, job title and major responsibilities, goals and tasks in relation to the product under consideration, environment, and often a quote that sums up what matters most to the persona when it comes to a particular product. The persona is used to help members of a development team understand the end user and keep that user in mind when product development decisions are made.

Pruit and Adlin [2] have described many advantages to using personas. Personas are role focused and put a human face on the research data. This personalization aspect makes it easy for a design and/or development team to identify with a prototypical user of their application. Furthermore, documenting all the personal details helps

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 313–320, 2011.

validate or invalidate the preconceived assumptions that the team has about the end users. Personas help the team stay focused on the aspects of the design that will satisfy the key target users [2]. Personas also provide a way to present functional requirements within the full context of the user [3]. Functional requirements are addressed in the persona by showcasing features and functions that users want or are currently using. Lastly, personas provide a detailed day-in-the life use case. These use cases make it easier for the product team to understand how real users would use their application to accomplish their daily work.

Most personas are limited in scope, focusing on the desktop user. This can be problematic since mobile technology has improved over the last few years and there has been a large increase in mobile devices and applications on the market [4, 5]. Many companies are utilizing this opportunity to go beyond desktop application development and are dedicating significant resources to creating mobile applications. Mobile devices and applications allow companies to target users that previously might not have been available to them in the desktop application market. This brings about new challenges because development teams do not know how to think about mobile users and have a vague notion of what is important. As we began the process of thinking about mobile users and how they work, we realized that the typical persona was insufficient for adequately describing them. There are significant differences between mobile users and desktop users that need to be incorporated into the persona. For example, mobile users perform tasks in short spurts and are often easily interrupted [6].

Personas have rarely addressed cultural differences in users because in many cases the user of the product was not perceived to vary across cultures, or more likely, because sales demographics did not indicate a significant target market to require a more cross-cultural representation. This assumption is not true for the mobile user. Mobile devices and applications extend the market place into regions previously not accounted for in desktop applications. National laws, cultural backgrounds, socio-economic backgrounds, religion, and language can all affect how people interact with a product between different countries [3]. These cultural differences in mobile behavior must be accounted for in the design of mobile products. The rise of global, mobile applications has created a new group of users that we are still learning about.

Companies are starting to conduct international ethnographic research to learn more about these mobile users. Ethnography is a process of gaining an understanding of work or activity as it occurs in its natural setting [7]. Ethnographic research that supports technology aims to inform product design and is conducted by shadowing mobile users throughout their day to understand their mobile tasks, environment, and the tools that they use to complete their tasks [8]. The rich data obtained from this research provides the material to create a robust set of personas.

Even though there is extensive literature on personas in general, there is very little information about how to create personas for mobile users and how to account for cultural differences in personas. This paper will present a case study of international mobile research that resulted in a set of rich, detailed, mobile personas. We will discuss the process that we used to create our set of personas including our mobile focus, how we utilized cross-cultural findings, and incorporated forward thinking sections to help foster innovation within our product development teams.

2 Methodology

The Oracle Mobile User Experience team conducted a large international ethnographic study on mobile users. Our research was conducted in Mumbai, Singapore, San Francisco, and New York. We followed 33 experienced mobile device users across seven business roles. In addition, we also administered questionnaires to 150 people on the streets.

Before we began the study, we identified several goals for our research. First, we wanted to observe where, when, and how mobile workers used their mobile devices. Second, we wanted to identify the types of content and applications that people wanted on their devices. Third, we wanted to observe the mobile work culture of each participant. This third aspect consisted of observing the participants' work-life balance, the percentage of time they spent on their mobiles, the time they spent working away from their desks, their etiquette and cultural practices, and the various ways they used their devices in social contexts. Lastly, we wanted to identify the prominent mobile user roles and create personas for each of them. Our research was based on the following four activities:

2.1 Observation

We accompanied each of our 33 participants one day for five hours, observing them as they went about their work and day-to day activities. We paid specific attention to mobile tasks, environmental changes and any tools (mobile devices, computers, and so on) that the participants used to complete their personal or work-related tasks. During the observation, we photographed and video-and-audio taped them completing their daily tasks, when appropriate.

2.2 Interviews

We spent two hours interviewing each of the 33 participants; one hour before the observation to learn about their backgrounds and work tasks and to build rapport, and one hour after the observation to follow up on our observations and to ask additional questions about their perceptions and usage of handheld devices.

2.3 Participant Photo Diary

Twenty-four of the 33 participants completed and returned a photo journal that required them to take pictures of their mobile activities based on tasks or environments that we specified. In addition, we asked participants to write about the tasks they performed in each location. For example they were asked to photograph and document the "nosiest" location where they typically use their mobile devices.

2.4 On-the-Street-Methods

We conducted other research while on the streets in each country to supplement the data that we collected during the follow-along observations. We administered more than 150 surveys while in the field, gathered print mobile advertising, captured pictures of other mobile advertising, and collected mobile artifacts. We visited a variety of mobile shops, including stalls on the street and high-end mobile stores. Researchers interviewed staff, gathered marketing materials, and learned how products and services were advertised and sold. The on-the-street research also allowed us to gather information about what types of features and devices buyers were requesting, and what some of their major purchasing considerations and pain points were.

3 Results and Discussion

After completing the research we analyzed the data. One of the key deliverables resulting from our analysis was a set of five user personas spanning four of our key business roles and one additional user persona representing the Gen Y population (the future users of our products).

Creating a persona template that conveyed all the information we wanted about the users was challenging. Our team had several brainstorming sessions and made multiple iterations of the template until we came up with a meaningful format. The following section describes how we addressed cross cultural issues, mobile users, and forward thinking aspects in our personas.

3.1 Personas for Mobile Users

Based on this research along with data from other participants we had previously run in mobile studies at Oracle labs, we knew the way a mobile user completes tasks differs significantly from the way a user performs similar tasks while sitting down in an office setting. Some examples of the mobile work style include: preference for short tasks, physically in movement, and frequently interrupted [6]. We did not want product teams to think it is acceptable to simply port over existing desktop applications to the mobile device without thinking about the mobile tasks at hand. We did the following to ensure that our personas focused on what users do while they are mobile:

1. We geared the title of the persona toward mobile usage. Each persona was called "Mobile Face" along with the title of the role. This title informed readers that the persona was focusing on the mobile aspects of a user's work.
2. We created a specific section in the template to describe mobile tasks. In this section, we described and illustrated both personal and work mobile tasks of the user.
3. We omitted information that was not mobile related. The only time we included non-mobile information was when we provided background information, such as demographics, describing work environments and occasionally to provide context in the day-in-life scenario. Focusing the personas on mobile information helped in the following ways:
    a. Reduced the size of the persona. The persona would have been too long and overwhelming if we included everything that a person did while using a desktop computer in the office.
    b. Focused product teams on mobile application tasks and eliminated inclusion of unnecessary desktop functionality.

4. We only included pictures in the personas that represented mobile users or aspects of their device interaction.

3.2 Personas for Cross Cultural Users

One of the goals in creating the personas was to highlight cross-cultural differences between mobile users. When we began the persona project we were not sure of the best way to accomplish this goal. We started by investigating the literature to determine how others were creating these types of personas. Unfortunately, we did not find a lot of literature on this topic. We found only one case study documenting how the Windows International Program Manager's team took their current personas and adapted them for a global marketplace [2].

Since we did not find many examples in the literature, our team brainstormed different ways we could integrate differences into our personas. We came up with the following three approaches:

1. Create a separate persona for each culture for every role.
2. Create a U.S. based persona for each role and include a section that documents the cultural differences found in the other two countries visited.
3. Create one persona for each role, but instead of having them all for the U.S., some of the roles would be characters based on the other countries visited. This approach meant that the identity of the persona must live in a city from that country, and that all references to places and customs would center on that same city and country. The cultural differences found during the research would be integrated throughout the persona without having a special section called out.

After conducting our research and analyzing the data, we discovered that the work tasks and flows of enterprise workers were very similar across the different cultures. However, there were observed differences between the cultures; a few examples include: which brand of mobile device they used, the type of transportation used, and the frequency of Short Message Service (SMS) usage. Overall, it didn't make good sense to create separate personas for each culture. While there were some cultural differences, there were not enough differences to warrant separate personas for each country and user role. We then debated on the other two options mentioned above.

We started out by creating U.S. based personas that included separate sections highlighting the cultural differences. When we presented these personas to other user experience professionals outside of our team they did not react favorably. They said that the cultural information seemed disjointed from the rest of the persona. They indicated that it conflicted with the goal of the persona, which was to get the reader focused on identifying with the character in the persona.

At this point we decided to switch approaches. We still created only one persona per role, but the roles were not all U.S. focused. We made sure to have at least one persona (within the set of personas) from Singapore and one from India, to remind product development teams that our products span across different countries and cultures. In addition, we made the decision to integrate any cultural differences into the personas without including a separate section for cultural differences. We did this by incorporating other findings from other countries in a realistic scenario for the given country of the persona under creation. For example, SMS was done frequently as a work task in Singapore and India, but not in the U.S. However, we still incorporated SMS into the U.S. personas because we knew this was something important to include in future Oracle applications. This approach was fairly easy to implement, and allowed for a single voice for each persona. In the end, although this approach increased the scope and diversity of the task flows it provided a cohesive way to include cultural differences in the personas.

3.3 Forward Thinking Personas

The mobile space is constantly changing with the advancements of technology. Therefore, we needed to make sure that our personas incorporated forward thinking information so that future products we build are not out-dated before reaching the market. We were able to incorporate forward thinking information into the personas in two different ways:

1. Most enterprise workers using mobile devices do not have specialized enterprise software installed on their devices. We observed mobile workers performing work-arounds by using their laptops, physical notebooks and calendars, or using pre-installed software that came on their device. We did not want our personas to show the current workarounds, but decided that it was important to streamline these tasks in the day-in-the- life scenario and show how they would optimally perform these tasks on their mobile device. We thought it was important for product teams to get a realistic picture of what users should be doing with their mobile device
2. We had a special section called "Mobile Future" which was dedicated to listing forward thinking requirements that we wanted development to consider in upcoming releases. These ideas were innovative concepts that rely on sensors, geocoding and technology standards that most products in the market do not currently support. Forward thinking concepts came from participants' comments on what they wish they could do or on experiences they've heard about or seen. We also inferred some concepts based on current limitations they faced or pain points they expressed. For example, a field service technician wanted to be able to connect his mobile device to a malfunctioning air conditioning unit to run diagnostics.

By paying special attention to these future needs, we helped development stay informed of new possibilities and areas to explore for future products. Also, it makes the personas relevant for a longer time and prevents them from becoming out-of-date quickly.

3.4 Mobile Personas

We generated five different personas that incorporated mobile, cross-cultural and forward thinking aspects (See Figure 1, for an example). Each persona had four tabs that described different aspects of mobile life. The first tab, "About Me" gave an overview of the persona including demographic information, personality characteristics, and technology usage (including mobile). The "Work" tab was a place to describe the person's job; including tools used, work places, and job goals. The job related information predominately focused around when the participant was mobile.

The third tab was called "Mobile Life" and provided details about personal and work mobile tasks along with a specific section on future mobile needs. The last tab was the "My Day" tab which was a detailed day in the life use case describing details of what the person does out in the field and how he or she interacts with her mobile device. Here is the list of the five personas that we created:

1. David Thompson: Field Service Technician, New York City, New York, USA
2. Amit Patel: Manager, Mumbai, India
3. Greg Roberts: Field Sales Representative, San Francisco, California, U.S.
4. Jing Tang: Retail Merchandiser, Singapore
5. Mary Turner: Generation Y, San Diego, California, U.S.

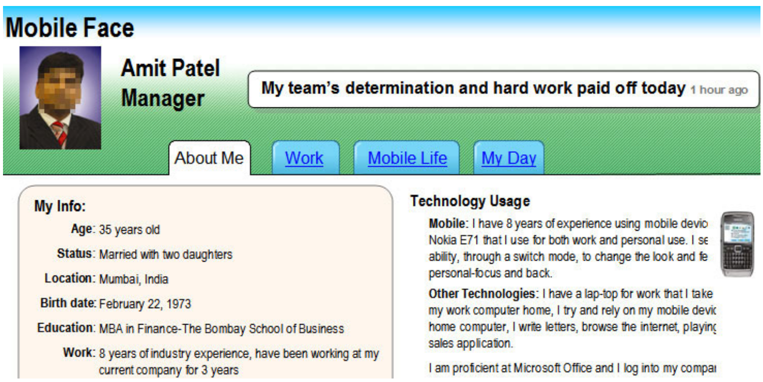


Fig. 1. Example Persona: Manager, Amit Patel

By creating our set of cross cultural mobile personas, we have established a new process when working with product development teams that keeps the focus on the mobile user and their environment. Developing mobile applications without the context of the day-to-day realities of the end user leads to designs that do not match the work style of today's mobile workforce. Previous to the introduction of the personas, product teams created features that were not mobile relevant and were better left to the desktop. Task flows were too long and layout was not simplified to the most important information. Phone capabilities (e.g. mapping addresses, auto-phone dialing, etc.) were not taken into account. With the help of the personas, we are better able to communicate the real needs of end users and deliver successful designs.

4 Conclusion

With the introduction of new technologies and with the shift towards globalization of products, the need to understand new users is increasing. Creating personas to help designers and product teams learn about end users has been used successfully for over a decade in the HCI community. However, traditional desktop personas must be modified to address cultural differences across users and to specify information unique to mobile users.

We provided a case study to show how personas can be geared toward both mobile and international users. When making mobile personas, consider including a section on mobile tasks, excluding a lot of information that is not related to mobile, and only include pictures that capture specific instances of mobile usage. As discussed, there are also different ways to include cultural differences in your persona. The first step is to evaluate how many cultural similarities and difference there are among your users. If there are many differences, consider creating individual personas for each country or region in your study. Otherwise, create a single persona for each role and either have a separate section to show the cultural differences or integrate the cultural findings into the persona without calling out each country where a finding was found. Lastly, don't forget to make your personas forward looking to inspire future designs. Include sections that address how users would ideally like to interact with their mobile device in your persona to help developers think through future use cases.

References

1. Cooper, A.: The Inmates are Running the Asylum. Macmillan, New York (1995)
2. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in mind throughout product design. Morgan Kaufmann, San Francisco (2006)
3. Quesenbery, W.: Making Personas Part of Your Team. Workshop Presented at User Friendly 2008, Shenzhen, China (October 2008), http://www.wqusability.com/handouts/personas-team-uf2008.pdf
4. Carolina, M.: Gartner Dataquest Forecast Analysis: Mobile Devices, Worldwide 2007-2014, 4Q10 Update (2010)
5. Gustav, S: Mobile Application Growth Surged in 2010-Distmo (2010), Nasdaq.com. http://www.nasdaq.com/aspx/company-news-story.aspx?storyid=201101101043dowjonesdjonline000212&title=mobile-application-growth-surged-in-2010-distimo
6. White, B., Rampoldi-Hnilo, L.: Design for the Mobile Experience (2009), http://usableapps.oracle.com/successStories/071115_mobileExperience.html
7. Martin, D., Sommerville, I.: Patterns of Cooperative Interaction: Linking Ethnomethodology and Design. ACM Trans. on Computer-Human Interaction 11(1), 59–85 (2004)
8. Blythe, S.: Fieldwork: Is it Product Work? Interactions (January/February 1997)

Motivating Change and Reducing Cost with the Discount Video Data Analysis Technique

Jody Wynn and Jeremiah D. Still

Missouri Western State University, Department of Psychology,
4525 Downs Dr., St. Joseph, MO 64507, USA
{jwynn,jstill2}@missouriwestern.edu

Abstract. Testing the usability of an interface is a critical phase of product development. However, it is often reported that analyzing the data from such testing consumes too many limited resources. We attempted to reduce this consumption by proposing a new technique, Discount Video Data Analysis (DVDA). We compared it with another popular accelerated analysis technique, Instant Data Analysis (IDA). Using IDA, evaluators analyze data after a series of usability tests, whereas DVDA calls for analyzing the data after every test in the series. Immediate analysis decreases the chance that subsequent test data will negatively interfere with evaluators' recall. Additionally, DVDA produces a video of the testing allowing the users' emotional responses (e.g., frustration) to be shared with developers who may be resistant to interface modifications. We found evaluators using DVDA identified more usability issues and provided more supportive evidence for each issue than evaluators using IDA.

Keywords: Data Analysis, Usability Evaluation, Discount Usability Testing.

1 Introduction

Two of the most prevalent usability evaluation methods used in the Human-Computer Interaction (HCI) community are expert reviews [1][2][3][4] and empirical tests [5][6][7][8]. The majority of studies comparing expert reviews and empirical tests indicate that the latter are superior for finding more severe interaction issues [9][10][11], but they are often discredited for being too time-consuming and expensive [4][9]. Ideally, HCI professionals need to have both the efficiency of expert reviews and the effectiveness of empirical tests. This need guided our development of the Discount Video Data Analysis (DVDA) technique.

The idea for DVDA was inspired by Kjeldskov, Skov and Stage's (2004) technique, which they coined Instant Data Analysis (IDA), of completing an empirical usability evaluation within a day [12]. The defining processes involved in IDA are to observe usability tests for a few hours then analyze and document usability issues later that day. Use of this procedure greatly reduces empirical testing costs compared with traditional methods.

Our DVDA technique diverges from IDA by stipulating that analyses occur after every test in a testing session rather than waiting until the end of the session. It has

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 321–328, 2011.

been shown that remembering new information can interfere with our ability to remember older information; especially when the information is highly similar [13]. Thus, when multiple tests are observed – as is done using IDA – subsequent tests could proactively interfere with the evaluators' recall of earlier usability issues. We predicted that documenting usability issues immediately following their occurrence ought to decrease memory errors.

Another difference between DVDA and IDA is that the former prescribes video recording the tests. These recordings may be employed to help resolve disagreements between study observers and to convey emotional aspects of the interaction – user frustration – to developers who might be resistant to changing an interface. Although the cost of video analysis may have seemed prohibitive in the past, the development of digital recording software makes it possible to record and edit videos in minutes rather than hours. To watch any portion of a video, the test evaluators can digitally "jump" to that point in the recording.

To examine if the differences between DVDA and IDA influenced the number of usability issues evaluators identified, both methods were used to analyze a commercial website. We predicted that evaluators using DVDA would identify more usability issues than evaluators using IDA. Additionally, we predicted that DVDA evaluators would produce more detailed reports without expending more resources than IDA evaluators.

2 Method

2.1 Participants

The university institutional review board approved all experimental procedures. Four undergraduate Introductory Psychology students were recruited to participate in a 20-minute usability test in exchange for course credit.

2.2 Facilitator and Evaluators

One Human Factors graduate student acted as the facilitator. Two Human Factors graduate students and two upper-level undergraduate Psychology students acted as usability test evaluators.

2.3 Apparatus and Stimuli

The usability tests were conducted in a private testing room equipped with a computer workstation. Participants interacted with a commercial news website. Remote observation took place in separate rooms, also equipped with computer workstations. Morae Usability Software [14] was used to record, remotely observe, and review the usability tests.

2.4 Procedure

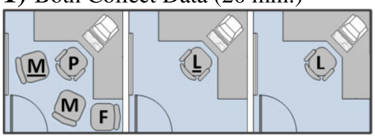
To compare DVDA to IDA, usability test evaluators used both techniques to analyze a website (see fig. 1). The evaluators were paired - one graduate student and one undergraduate student. Each pair employed either the IDA or DVDA technique. The DVDA evaluators followed instructions for the DVDA methodology (see Appendix A). The IDA evaluators followed instructions for the IDA methodology [12]. Both sets of evaluators followed our standard documentation for severity ratings (see Appendix B).

Both Techniques	**DVDA Technique**	**IDA Technique**
P: Participant	L: Data Logger	L: Data Logger
F: Facilitator	M: Monitor	M: Monitor

Layout

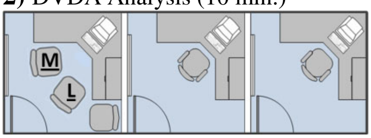
Outline

1) Both Collect Data (20 min.)



- Facilitator read tasks
- Monitors watched for errors
- Data Loggers remotely watched for errors

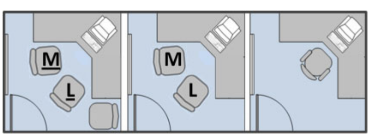
2) DVDA Analysis (10 min.)



- DVDA evaluators analyzed data
- IDA evaluators did not meet or discuss the test

- Repeated steps (**1**) and (**2**) for the second, third, and fourth test
- Monitors and Data Loggers switched roles before each test

3) IDA Analysis (40 min.)
Both Create Reports (20 min.)



- IDA evaluators analyzed the data
- DVDA evaluators created a report
- IDA evaluators created a report

Fig. 1. The left column shows the room layouts with key actors. The right column outlines the experimental procedure used to compare the DVDA technique with the IDA technique. It is important to note that both accelerated usability techniques took the same amount of time. The DVDA evaluators spent 10 minutes after each of the four tests to perform their analysis; resulting in 40 minutes of analysis. IDA evaluators spent 40 minutes performing their analysis after all the tests were complete. They both took an additional 20 minutes to create their data analysis reports.

A facilitator acted as a liaison between the evaluators and the usability test participants. She instructed the participants to use the think-aloud usability test protocol [8], asking them to say what they are thinking, doing, and feeling while interacting with the website.

One test evaluator for each methodology sat in the room with the test participant and performed the role of test monitor. The monitors did not take notes, but the DVDA monitor wrote down the elapsed time of the recording when she recognized a usability issue. Sometimes she wrote a letter or word by the time notation to remind her of what the issue was (extensive notes were not taken).

The other two evaluators acted as data loggers and sat in separate rooms. They watched the test on their computer screens and took extensive notes. The DVDA data logger made a notation of the elapsed time of the recording each time he documented an issue.

During the 20 minutes allotted for each test, the facilitator instructed the tasks to be performed by the participant (e.g., search for grocery coupons, subscribe to a channel, place a furniture ad, etc.). At the conclusion of each test, the DVDA evaluators took ten minutes to analyze the test they had just observed (following the protocol outlined in Appendix A). During their analysis, the DVDA evaluators referred to the data logger's notes, the website, and the video of the usability test to aid their recall of the issues encountered by the participant. After ten minutes, the data loggers and test monitors switched roles and took their respective places in the testing room and observation rooms.

After the fourth and final 20-minute test, the DVDA test monitor and data logger spent 10 minutes on their last test analysis and 20 minutes creating a report. At this time, the IDA test monitor and data logger met for 60 minutes and analyzed the usability test and created their report.

All four evaluators finalized their data analyses, determined the severity of each issue, and typed their report within 60 minutes. The DVDA evaluators performed their 40-minute analysis in four 10-minute increments after each test, and then spent 20 minutes producing their report after the entire usability testing session was done. The IDA evaluators performed their 40-minute analysis after watching all four tests. Directly following their analysis they spent another 20 minutes creating their report. Once both teams completed their usability reports they discussed their error severity categorization. We hoped this would promote interobserver reliability across the two teams. Each team reviewed the other team's issue descriptions and agreed on the severity rating of issues they found in common.

The IDA report only included a description and severity rating of each issue. However, the DVDA report included a description, severity rating of each issue, the number of test participants who experienced each issue, and the time each issue occurred.

The IDA and DVDA techniques' performance were compared in terms of quantitative and qualitative measures. The quantitative data extracted from the reports were the total number of unique issues in each severity category, the number of issues in each category identified using DVDA and IDA, and the number of test participants that had experienced each problem. The qualitative measure was the thoroughness of the issue descriptions.

3 Results

Out of thirteen total issues, twelve were reported by the DVDA evaluators and ten by the IDA evaluators. The results for both methods were similar, the only differences being in the "Serious" category. In that category, the DVDA evaluators discovered three issues that the IDA evaluators did not; conversely the IDA evaluators discovered one issue that the DVDA evaluators did not.

There was a qualitative difference in the content of the usability reports. The DVDA reports were more detailed than the IDA reports and included the number of users experiencing each problem. The DVDA reports also included a reference to the supporting videos via a timestamp on each issue.

4 Discussion

As expected, more usability issues were identified using the DVDA method than the IDA method. The key quantitative difference was that the DVDA method captured the number of test participants that experienced each issue. Information regarding frequency and severity of interaction issues is useful for prioritizing issues for redesign or reengineering [15]. Knowing which participants experienced each issue also allows the evaluators to identify outliers – those participants who show a trend for encountering more issues than other participants. Further, DVDA evaluators were able to include more descriptive elements in their report, and a video recording of each test was produced to support this information. The DVDA evaluators were able to explain the issues, indicate how many users had experienced the issue, and where the issue could be found in the video. Having the ability to reference the videos may become important if the evaluators need to justify the validity of the issues they report to developers.

These additional pieces of evidence are valuable for usability professionals, especially those attempting to gain credibility and respect from product developers. Nielsen [16] says, "As [a company's] usability approach matures, organizations typically progress through the same sequence of stages, from initial hostility to widespread reliance on user research." It is common for usability professionals to find themselves in a hostile environment where user-centered design and usability testing are viewed only as activities that create more work for developers. The mindset of the developers can be the greatest obstacle for deploying usability practices within an organization [17]. An effective way to convince developers of the importance of usability testing is to have them watch customers using their products [6]. When developers cannot attend the testing session, the video recording included in DVDA allows them to see users struggle with problematic design elements.

Although this study suggests that DVDA is a valuable technique there are some limitations. Due to time restrictions, only four 20-minute tests could be performed in this study. The evaluators only had time to find thirteen issues in total, but the DVDA technique revealed two more issues than the IDA technique.

We suggest a follow-up study in which the usability testing session is four or five hours long as prescribed by the IDA methodology. We would expect that given a longer delay between the experience and when it is retrieved from memory, significantly more usability errors will be forgotten by professionals using the IDA than the DVDA technique. Since the DVDA technique allows evaluators to record their insights when they are freshest in their memories, recalling the events of each test cannot be interfered with by observing other events in subsequent tests.

Both DVDA and IDA methods allow researchers to finish their analysis in a fraction of the time it would take using traditional methods. This study suggests that DVDA is capable of revealing more usability issues and producing a richer report than the established IDA technique.

Acknowledgments. We thank Sujoy Kumar Chowdhury, Joseph Grgic, Deanna Knox and Jacob Wharton for performing the usability evaluations.

References

1. John, B.E., Packer, H.: Learning and using the cognitive walkthrough method: A case study approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 429–436 (1995)
2. Frøkjær, E., Hornbæk, K.: Two psychology-based usability inspection techniques studied in a diary experiment. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction, pp. 3–12 (2004)
3. Muller, M.J., Matheson, L., Page, C., Gallup, R.: Methods & tools: Participatory heuristic evaluation. Interactions 5(5), 13–18 (1998)
4. Nielsen, J.: Usability inspection methods. In: Conference Companion on Human Factors in Computing Systems, pp. 413–414 (1994)
5. Dumas, J.S., Redish, J.C.: A Practical Guide to Usability Testing (Rev Sub.). Intellect Ltd., Bristol (1999)
6. Dumas, J.S.: Stimulating change through usability testing. SIGCHI Bulletin 21(1), 37–44 (1989)
7. Frøkjær, E., Hornbæk, K.: Cooperative usability testing: Complementing usability tests with user-supported interpretation sessions. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1383–1386 (2005)
8. Rubin, J.: Handbook of usability testing: How to plan, design, and conduct effective tests. John Wiley & Sons, Inc., Chichester (1994)
9. Jeffries, R., Desurvire, H.: Usability testing vs. heuristic evaluation: Was there a contest? SIGCHI Bulletin 24(4), 39–41 (1992)
10. Karat, C., Campbell, R., Fiegel, T.: Comparison of empirical testing and walkthrough methods in user interface evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 397–404 (1992)
11. Thyvalikakath, T., Monaco, V., Thambuganipalle, H., Schleyer, T.: Comparative study of heuristic evaluation and usability testing methods. Studies in Health Technology and Informatics 143, 322–327 (2009)
12. Kjeldskov, J., Skov, M.B., Stage, J.: Instant data analysis: Conducting usability evaluations in a day. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction, pp. 233–240 (2004)
13. Dewar, M.T., Cowan, N., Sala, S.D.: Forgetting due to retroactive interference: A fusion of Müller and Pilzecker's early insights into everyday forgetting and recent research on anterograde amnesia. Cortex 43(5), 616–634 (1900)
14. TechSmith Corporation, ©1995-2010, `http://www.techsmith.com/morae.asp`
15. Woolrych, A., Cockton, G.: Why and when five test users aren't enough. In: Proceedings of IHM-HCI 2001, vol. 2, pp. 105–108 (2001)
16. Nielsen, J.: Corporate usability maturity (2006), `http://www.useit.com/alertbox/maturity.html`

17. Bak, J.O., Nguyen, K., Risgaard, P., Stage, J.: Obstacles to usability evaluation in practice: A survey of software development organizations. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges, pp. 23–32 (2008)

Appendix A: Discount Video Data Analysis (DVDA) Procedure

DVDA may be used with any empirical usability testing method which can be recorded via screen-capture software. The mechanics of the testing technique are not constrained by DVDA, although the following procedures must be included in the usability testing phase:

1. Video record the test using screen-capture software.
2. One evaluator act as test monitor and sit in the testing room with the test participant. Interact with the participant according to the testing method being used. Focus on the participant and do not take extensive notes. Notate the time each issue occurs, along with a letter or word to aid in recalling the issue later. If a subsequent participant encounters the same issue, put a tick mark beside the original notation to indicate replication of the problem.
3. One or more evaluators act as data logger(s) and take extensive notes, including the time each issue occurs. If a subsequent participant encounters the same issue, put a tick mark beside the original notation to indicate replication of the problem. Focus on documenting issues and do not interact with the test participant.
4. Evaluators must use a common notation time (e.g., recording elapsed time).

Data analysis is performed immediately after each test using the following procedure:

1. All evaluators meet at the conclusion of each test for approximately ½ the time the test took to collect the data. For example, if the test data collection took 60 minutes, spend no more than 30 minutes on the analysis.
2. Compare time notations, discuss experience, and document common issues.
3. For issues not found in common, explain the issue to the evaluator(s) who did not see it. If the evaluators agree it is a problem, document it.
4. To review any part of the test, "jump" (i.e., digitally skip) to that event in the video.
5. Count the number of tick marks by each issue and add one for the original issue to tally how many times each problem was experienced by a test participant. Include this information in the usability report.
6. Make screen shots if needed to augment the documentation.
7. Make video clips if needed to augment the documentation.
8. Categorize the issues based on severity criteria. See Appendix B for example severity criteria.
9. Perform a final review and compile a usability issue report that includes a description, the severity, and the number of participants who had each issue.

Appendix B: Usability Issue Severity Criteria

Critical: They could not reach the intended goal. This could be due to an error or failure in the website or because the path to complete the task was obscure.

Serious: They took an excessive amount of time to reach the intended goal. This could be because the path to complete the task was not obvious.

Aesthetic: They could complete the intended goal but were not pleased with an aspect of the look and/or interaction.

What You See Is What You Don't Get: Addressing Implications of Information Technology through Design Fiction

Ludwig Zeller

Dept. Design Interactions, Royal College of Art (RCA),
Kensington Gore, SW7 2EU, London, United Kingdom
ludwig.zeller@network.rca.ac.uk

Abstract. This paper outlines three design projects that address the implications of current and emerging information technologies for the interests, abilities and psychological condition of the people who use them. It is analysed how these projects address the emerging needs of the "digital natives": the generation of young people that grows up with digital information technology from an early age on. The origin and usage of the term "design fiction" is explained and a comparison with science fiction is put forth with a special focus on the work within the department. It is shown how design fiction can be used as a "Trojan horse" for communicating unconventional and unforeseen implications to a larger audience.

Keywords: design methods, information technology, design fiction.

1 Introduction

This paper places emphasis on a special kind of "design fiction", which sees human nature as much more complex than the concept of a "user" in classical UX research. By using fictive products and design scenarios as a medium, the projects in this case study visualise mindsets of future generations that could emerge from the information technology we start to live with today.

While the following three project examples[1] are taken from our work at the Department of Design Interactions, their interpretation has been written only by the author of this paper.

2 Design for the "Digital Natives"

The generation that is growing up in the networked world, which we are living in for a short time now, can be called the "digital natives" [4]. These young people do not remember the difference between analogue and digital, since they never experienced the technological paradigms that have been around before them.

[1] Special thanks to Gerard Rallo, Fiona Raby and Anthony Dunne for providing material about their projects as well as Steffen Fiedler for providing feedback.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 329–336, 2011.

Fig. 1. Which new kinds of social relationships could emerge through the HCI we create today? “Conversations Challenger” by Gerard Rallo

Many of these individuals are very talented with technology and approach it intuitively even though they do not necessarily know much about how these machines work internally. They are also very multi-task capable and seem to be able to deal with a high amount of information input, while some are also easily distracted and may have minimal direct social contact with each other. Social interaction through networked messaging and friend platforms play an important role in their coming of age process.

The design of human-computer interaction will have to adapt to new kinds of personal interests. What kind of traits will develop through the on-demand access to an overflowing yet personalised amount of visual and auditive media, which is immediately shareable with networked friends? How will their sense of identity balance between virtual and physical space?

Even though it is obviously hard to forecast the future and all consequences these digital innovations will have on forthcoming generations, design fiction offers an approach to visualise cultural implications and speculations of the interactions between technology and society and thus offers an entry point for discussion.

3 Examples of Design Interactions Projects

In the following several projects from the Department of Design Interactions at the Royal College of Art London will be illuminated. It has to be noted that these only represent a subset of the activities that are taking place at the department, but have been selected to illustrate a spread of approaches towards the specific topic of this paper. Please note that different designers have developed these projects independently from each other at different times. The "dromolux" project, which is presented last, is currently in development by the author of this paper. Therefore, more detailed information will be given about that specific project.

The "Conversations Challenger" [5] by Gerard Rallo for example is part of the series "Devices for Mindless Communication". The object in fig. 1 is a display device with an attached microphone. Being placed between two talking persons it is designed to detect the topic of the conversation and give meaningful search results from the Internet as a constant, contextual augmentation for the owner.

This fictive product explores interactions by putting our desire for time-efficient, partly automated communication to an extreme. Many innovations are aimed to minimise the cognitive impact of unpleasant activities. How far can this be taken in human interaction?

Another example for our research about future ways of visual perception of information is the "S.O.C.D." [1] project by Anthony Dunne and Fiona Raby. Sexual Obsession Compulsive Disorder (S.O.C.D) is a common behavioural phenomenon. It describes involuntary, sexually orientated fantasies that cause a spontaneous feeling of shame. Dunne and Raby explore a variation of this disorder in which somebody would like to consume pornographic media, but feels guilty when doing so. The "S.O.C.D." device in fig. 2 is a setup of a DVD player, monitor and a long, rubbery sensing device. The viewer has to hold his hand onto the sensor in order to play the video clip. When it detects the obsessive arousal of the viewer, the clip is increasingly distorted through pixelation. This helps the viewer to learn how to maintain a balanced state while watching.

Instead of trying to ignore or even cure the "perverted" or "ill" behaviour, this design proposal by Dunne and Raby is specifically adapted "for" the special normality that people with S.O.C.D. are living in. This project is situated within their "Do You Want to Replace the Existing Normal" series, which looks beyond classical model relationships between users and machines and reflects a more fine-grained perspective on what it means to be human.

In fig. 3 the "dromolux" [7] project by Ludwig Zeller is shown as another design example that addresses the possible increase in cognitive performance of future generations. This device is a speed-reading trainer that flashes words at an increasing pace using a strobe light and very short exposure. Used together with appropriate medication it could help elderly or handicapped people to fight their cognitive flaws.

This project originated out of a set brief to visualize 120 seconds of future. The basic supposition was that in future the amount of information within a given time span will in general be higher than today. Even though that seems to be an obvious assumption a number of research hints could be found that undermine this statement. Palfray and Gasser for instance report that the amount of digitally created, stored and replicated content has been 161 billion Gigabytes in 2007, while it was estimated to rise to 988 billion Gigabytes already in 2010 [4]. While this high growth seems to be sustaining, they also state that the natural human ability to process information stays comparatively slow with around 126 bits per second.

Fig. 2. Automatic pixelation for regulated media consumption. ”S.O.C.D” (Sexual Obsessive Compulsive Disorder) by Anthony Dunne and Fiona Raby. (Image courtesy of Francis Ware).

But for the members of our information society the ability to process a growing amount of data is a necessity or at least an advantage. Therefore, evolutionary processes and improvements in the development of so-called smart drugs are likely to raise the pace of our perception. In 2008 the “Nature” magazine conveyed an online survey of cognitive enhancement drug use amongst its readers [8]. Approx. 20% of the 1400 participants reported having tried drugs for brain enhancement with methylphenidate (“Ritalin”) as the most popular followed by “Modafinil”. These prescription drugs are known for their wakefulness and concentration promoting effect and used especially in the treatment of attention deficit hyperactivity disorder (ADHD). Ethical questions arise from this report regarding the usage of doping in general market competition, the unforeseen side effects and disadvantages to people without access.

As part of the project these and other research findings have been extrapolated and translated into a design fiction proposal that sits within a future scenario. The “dromolux” object is designed for a world in which the global processing of information has grown enormously and the medication with brain enhancement drugs became morally accepted, more effective and safe.

In a society predicated on the transfer and consumption of information, the wish for never-ending cognitive function could replace the desire for endless healthiness and beauty. How will the digital natives deal with their mental decay in a few decades time? The "dromolux" would be used at nursery homes where elderly people could opt-in for a speed-reading training scheme. At the time of this writing a working prototype has been constructed and it is planned to visualise its usage in a video scenario with cast actors.

What other kinds of services could emerge together with future developments in smart drug medication? Along with these developments, new ways of viewing and processing of information come up that challenge our familiar metaphors of screening, depiction and watching.

4 Design Fiction as an Alternative UX Research Method

This section will elaborate on the notion of design fiction as a method for interaction design. The term itself has probably been coined by science fiction author Bruce Sterling in 2009 [9]. Since then a number of reports and essays dealing with its approach and impact have been published on the web, but little citable information is available to date as its precise definition and differentiation to other approaches seems to be still under negotiation between practitioners of the field.

In general design fiction creates visualizations of not yet commonplace constellations of technology and society through design. But this strategy of exemplifying a design proposal in the context of a scenario is also frequently used in other disciplines such as industrial design. As it would be beyond the scope of this paper to discuss these positions as well, this text will have to focus on a sub-set of approaches that are practiced at the Department of Design Interactions at the Royal College of Art in London.

The literary discipline that the name "design fiction" is derived from is "science fiction". Science fiction as a popular genre in cinema and writing is commonly known for its exaggerated perspective on alternative or plausible worlds. But this genre has had a number of great successes in communicating science and technology to broader audiences.

Visualisations in titles such as "Minority Report" (2002) have strongly coined the expectation of what the future of interaction will look like. Julian Bleecker from the "Near Future Laboratory" looks at how the public and the media mix scientific facts with science fiction by analyzing Google search results for "Minority Report Interface": "Search results present us with stills from the film right next to things that look like DIY garage science projects, to demonstrations of touch panels at industry trade shows like TED and CeBIT, to reviews of the iPhone interface using *Minority Report* as a point of reference, to promises that the *Minority Report* interface is just around the corner — wherever that corner may be." [10] Fictional literature and the design of actual technology engage in a mutual, cultural exchange. This goes along with the research that David Kirby is putting forth in his publications for several years now [2][11]. The visualizations of cinema pave the paths for creators who develop technologies by opening up more access to funding and dissemination of their according work on the one hand, while emerging technologies inspire film makers on the other.

Fig. 3 Never ending cognitive function replaces desires for healthiness and beauty. The "dromolux" flash reading machine by Ludwig Zeller.

The design fiction that is created at the Department of Design Interactions also uses designed visualisations to allow for the imagination of future or alternative scenarios. But contrary to science fiction the goal of the department is less to serve common mainstream interests and beliefs of future technology, but to open up discussions that could allow rethinking these conventions. Futurologist Stuart Candy who is a visiting lecturer at the department describes the intended outcome to be "killer implications" in contrast to the traditionally strived for "killer applications" [12]. By giving theoretical ideas and speculations a tangible form, they create a direct awareness and empathy for otherwise incomprehensible needs and situations.

While there is not one specific design method, a certain philosophy of approaching a project can be outlined. Often the analysis of emerging social phenomena or technologies stands in the beginning of a design fiction project. In direct interaction with researchers a basic understanding from the perspective of an outsider is created.

Afterwards, the search for implications of the given situation can be inspired by a number of even exotic sources such as newspaper reports about unforeseen interactions with similar, already existing technologies. Design Interactions guest lecturer Tobie Kerridge has documented how he and his collaborators use a spectrum of methods in order to broaden the research material and engage with public and science from an early moment on [13]. The list of techniques for public engagement ranges from surveys, workshops, interviews, presentations and debates to artistic interventions. The experiences from this research will then be accounted for in the creation of a new design proposal.

A common way of framing an implication is the "Trojan horse" strategy as illustrated in fig. 4: The design outcome keeps the same high quality standard that its according discipline is used to in order to maintain its believability and seriousness.

But this familiarity to the audience is then used to open up a subtle channel of believe for reporting a set of unconventional ideas and implications. This strategy follows Samuel T. Coleridge's idea of the "suspension of disbelief" that he originally put forth as a writing technique for fictional literature.

Accordingly the following dissemination of a project can happen through a number of various occasions and channels such as design symposia, conferences, lectures, exhibitions and the art world. An interesting way of presenting are "performative lectures" that a number of projects made use of. In this case the design fiction project is not only presented within the own design community but also presented to the public or special scientific and administrative audiences at according events. It is interesting to see that for these presentations the suspension of disbelief has to be sustained beyond the actual design process up to the point where the presenting designer fits herself in the role as a member of the given community. From experience it can be stated that these kinds of intense, direct interactions with external audiences have brought up very interesting opportunities and project environments for the involved designers.

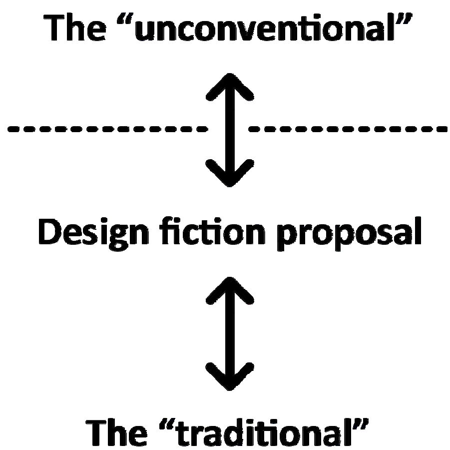


Fig. 4. Design fiction as a "Trojan horse" strategy. An intriguing design proposal within the realm of the thinkable is used as an interface to a hidden implication.

5 Conclusion

Fictive depictions are a powerful medium for shaping our self-conception in relation to our technology. They allow for a presentation that can give insights about the implications of the introduction and invention of technology.

For this purpose a scientific correctness is not completely necessary, it is more important to outline telling, daily-life situations that emerge from the combination of man and machine. Langdon Winner states that "[k]nowing how automobiles are made, how they operate, and how they are used and knowing about traffic laws and urban transportation policies does little to help us understand how automobiles affect the texture of modern life. In such cases a strictly instrumental/functional understanding fails us badly."[6]

As interaction designers we have to keep in mind the blurry and unforeseen consequences of our products within the private sphere of the people who use them. “We tinker with your philosophy by direct manipulation of your cognitive experience, not indirectly through argument.” [3] Our creations will have a direct impact for the self-conception of many people that live with them.

Design fiction does not illustrate the status quo, but illustrates what a life with emerging technologies could be like - for better or worse. This is a kind of user experience research that outlines consequences and implications that could not be highlighted in the traditional terms of efficiency. People are not seen as a “user” entity within a larger system architecture of interaction, instead it is tried to stay aware of all the hidden nuances and traits that a human being can have.

In this paper some traits of the emerging group of “digital natives” have been portrayed shortly in regards to skills and social interactions. Questions about yet unknown effects of growing up with ubiquitous information technology have been asked.

In relation to that, three project examples from the Design Interactions Department of the Royal College of Art London have been given for a kind of design that addresses future and speculative developments. These projects use a multitude of materials and media in order to tell intriguing stories about the present and future relationship between society and technology.

Furthermore this kind of design fiction was compared to mainstream science fiction depictions. Fictive illustrations were described as a general, powerful tool for shaping a popular understanding and vision about our life with technology. Eventually design fiction was presented as an alternative perspective on user experience research that takes complex consequences and unforeseen implications of the interactions between system and “user” into account.

References

1. Dunne, A., Raby, F.: Portfolio, `http://www.dunneandraby.co.uk`
2. Kirby, D.: Cinematic Science. In: Bucchi, M., Trench, B. (eds.) Handbook of Public Communication of Science and Technology, pp. 41–56. Routledge, New York (2008)
3. Lanier, J.: You Are Not A Gadget. A Manifesto. Borzoi Books, New York (2010)
4. Palfray, J., Gasser, U.: Born digital: understanding the first generation of digital natives. Basic Books, New York (2008)
5. Rallo, G.: RCA Summer Show, `http://www.di10.rca.ac.uk/gerardrallo`
6. Winner, L.: The Whale and the Reactor: A Search for Limits in an Age of High Technology. University of Chicago Press, Chicago (1988)
7. Zeller, L.: Portfolio, `http://www.ludwigzeller.de`
8. Maher, B.: Poll results: look who’s doping. Nature 452, 674–675 (2008)
9. Sterling, B.: Design fiction. Interactions XVI(3) (2009)
10. Bleecker, J.: Design Fiction: a short essay on design, science, fact and fiction, Near Future Laboratory (2009), `http://www.nearfuturelaboratory.com`
11. Kirby, D.: Lab Coats in Hollywood. Science, Scientists, and Cinema. MIT Press, Cambridge (2011)
12. Candy, S.: Killer imps, blog entry (2009), `http://futuryst.blogspot.com/2009/03/killer-imps.html` (accessed on 2011-02-24)
13. Beaver, J., Kerridge, T., Pennington, S.: Material beliefs. Interaction Research Studio, Goldsmiths. University of London, London (2009)

Modeling Users' Data Usage Experiences from Scientific Literature

Jian Zhang[1], Chaomei Chen[1], and Michael S. Vogeley[2]

[1]College of Information Science and Technology, Drexel University
[2]Department of Physics, Drexel University
{jz85,cc345,msv23}@drexel.edu

Abstract. In the new data-intensive science paradigm, data infrastructures have been designed and built to collect, archive, publish, and analyze scientific data for a variety of users. Little attention, however, has been paid to users of these data infrastructures. This study endeavors to improve our understanding of these users' data usage models through a content analysis of publications related to a frequently cited project in the data-intensive science, Sloan Digital Sky Survey (SDSS). We find that 1) Content analysis of scientific publications could be a complementary method for researchers in HCI community; 2) although SDSS produced a large volume of astronomical data, users did not fully utilize these data; 3) users are not only consumers of scientific data, they are also producers; and 4) studies that can use multiple large scale data sources are relatively rare. Issues of data provenance and usability may prevent researchers from doing research that combines such data sources. Further HCI study of detailed usability issues associated with data infrastructures in the new paradigm is eagerly needed.

Keywords: User Modeling, Data usage, Sloan Digital Sky Survey, Usability.

1 Introduction

Science is stepping into its "fourth paradigm" of data-intensive science. In the new paradigm, data infrastructures play a crucial role in capturing, preserving, publishing, and analyzing scientific data [1]. Data infrastructures, particularly data access portals that are mainly in the form of Web sites, have become information-intensive applications.

In the new paradigm, most existing studies are focused on techniques and tools for building up infrastructures to provide users with data at any time, any place, and in any form. However, relatively little attention has been paid to problems associated with human-computer interaction (HCI) perspective. For example, designers of these data infrastructures, particularly data access portals, are usually domain experts, instead of HCI experts. Also initial proposals of these infrastructures had their specific research purposes, which normally formed the tasks that data infrastructures were designed for at design time, but not tasks performed by real users at use time[2].

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 337–346, 2011.

To better tackle these HCI problems, it is necessary to first study how users actually used scientific data [3, 4]. This study answers this call and endeavors to model users of scientific data, particularly, their data usage behaviors. Different from normal studies of user modeling in HCI, which focus on modeling users behaviors based on one system's functions[2], this study looks at more general user behaviors that are not associated with detailed functions in a data infrastructure. We believe these general user models could become the foundations for specific designs.

The new science paradigm happens in various disciplines. For our work we chose a highly cited data-intensive project [3], the Sloan Digital Sky Survey (SDSS), as our case study. SDSS is considered an early milestone of the fourth paradigm[4]. It aims to map the large-scale structure of the Universe seen in the distribution of stars, galaxies, and quasars. The project collected digital images of 300 million astronomical objects, and digital spectral of about 1.3 million objects[5]. This data is publicly accessible on Skysever (cas.sdss.org), which is information-rich and -intensive.

2 Related Work

The Fourth Paradigm: Data-intensive Science

Jim Gray envisioned that science is stepping into its fourth paradigm[1]. Empirical science appeared a thousand years ago, theoretical science appeared a few hundred years ago, and computational science appeared just a few decades ago. Today a new paradigm is emerging, data-intensive science[6]. In the era of data-intensive science, the majority of scientific data will be captured by advanced sensors and instruments on a 24/7 basis, or be generated from scientific simulations[7, 8]. Soon scientific data archives will become new publishers of scientific artifacts, including not only scientific data, but also documents, records, and publications [3, 8].

While many challenges arise in implementing the technological infrastructure in the new paradigm, those who envisioned this paradigm have recognized that more challenging issues come from understanding of their users. Gray[9] indicated that "behind even the most difficult technical problems lies an even more fundamental problem: assuring the integration of the cyber-infrastructure into human workflows and practices."(p, 198). Borgman [3] emphasized that "[B]uilding a technical framework for scholarship is much easier than understanding what to build, for whom, for what purposes, and how their usage of the technologies will evolve over time."(p, 3).

Studies for modeling scientific data users

Existing studies of scientific data users in the new paradigm have identified and modeled users' behavior ranging from data collection, to data curation, data sharing, data publishing, and data access.

For collecting scientific data, [7] revealed that scientific data could come from observations, simulations, or both. The volume of scientific data from all of these sources keeps increasing, and most data is born digital [7]. After being captured, scientific data can be maintained in a variety of locations and data structures, like 1) Informal local data, 2) Structured local data, 3) Informally published data, 4) Community published data, 5) New shared repositories, 6) Reference repositories, and 7) Federated reference repositories, as defined in [4]. In general, data in the seven categories could fall into three domains: private data domain, community sharing domain, and public domain [10]. Traditionally, "big science" fields like astronomy have more data in the public domain. But "small science" fields also are experiencing a data deluge, and many of these fields demand data publishing for reuse by others[8]. When scientific data enter the public domain and are used primarily by those who are not the data collectors, data access behaviors become important to the data system designers and managers. Existing studies of data access behavior have focused on examination of the users' access logs for system optimization, load balance, or other management purposes[11] [12].

Studies cited above have revealed insights about the users of scientific data in the new paradigm. Researchers collected information regarding "what-they-want" and "how-they-access". However, little is known about how scientific data are actually used, particularly in the scenario of public data domain, where the data collectors, system designers and managers are a small fraction of final data users.

Research questions
Within the scope of SDSS project, this work focuses on the following research questions.

1. SDSS produced a data archive that contains millions of astronomical objects. How many objects were actually used in SDSS-related studies? Did scientists take advantage of the large volume of SDSS data and use a large number of objects?
2. Besides SDSS data, what are other data that were co-used with SDSS data? What are the co-usage patterns?

3 Methods and Materials

Survey and interview are two of the most common research methods for studying users' information behavior [13]. It is difficult, however, to recruit enough participants for surveys or interviews in this study because most of the targeted scientists are in external institutes. Therefore we turned to scientific publications, which in general clearly record scientists' data usage information. We used the content analysis method[14] to reveal scientists' data usage behavior from SDSS-related publications. Many previous content analysis studies have proven that this unobtrusive method can reveal insights with less effects of bias caused by participants in surveys or interviews. In addition, we conducted a small focus group study with three astronomers in our institution (led by the third author of this article) for development of the code book for this content analysis.

3.1 Contents of Code Book

Following the procedures of content analysis study [14], we designed a code book that contains concepts and process of this content analysis. The rest of this section highlights some key parts of the code book for readers to easily understand results presented in the following sections. For brevity, this paper only lists contents related to this study. The full contents can be found online[1].

Basic concepts

The key concepts used related to this study include the "unit of analysis", "data", "data source", and "object".

The *Unit of analysis* in this study is an SDSS-related paper, including its full-text contents and associated bibliographic features.

Data in this analysis means observational data collected by astronomical devices (telescopes on the ground and in space).

A *Data source* refers to a unique set of data collected by various astronomical projects and studies, such as the SDSS or Asteroid Database.

An *object* means an astronomical item with certain properties, like asteroids, stars, galaxies. One object could contain other objects of a different type, e.g. a galaxy with millions of stars. Therefore when counting the number of objects used in a study, the object refers to the major item that a study analyzed. For instance, if a paper focused on galaxies, the number of galaxies is the number of objects, rather than the number of stars in those galaxies.

Coding process

The process of coding one paper includes three steps. In step one, coders read the paper's title and abstract in detail to learn the general context of the paper. During step two, coders may skim over the coded paper, but need to read in detail the "introduction" section and "sample" or "data" or "method" sections to have confidence of coding the paper. Once finished reading, in the final step, coders start to code the paper by answering a set of questions and recording their answers. During the coding, they may refer to the paper again if necessary.

Major questions that coders need to answer

A set of questions that could reveal scientists' data usage behavior was designed for coders to answer from the set of papers. Coders first need to identify the type of papers from six categories that were developed from our focus group interview: "Theoretical paper," "Discovery paper," "Analysis paper," "Method paper," "Observational paper," and "Others."

For each paper, coders need to judge if the paper DIRECTLY used observational data (i.e. they analyzed the data themselves rather than simply refer to earlier analyses). If used, coders need to identify how many data sources were used and what they were. The name of a data source ordinarily appears in the "introduction" or "data" sections of a paper. For data that were collected by the authors for their studies, we named the data source as "Their own data collection". In the case of data sources analyzed in previous studies and referenced in the current paper, we named the data source as "Referred to previous studies".

For each data source identified in a paper, coders need to find the number of objects used in the data source. If there is no such information available on a paper, coders will label as "No descriptions". For each data source, coders need to find out the number of objects that were finally analyzed in the coded paper.

[1] http://nevac.ischool.drexel.edu/~james/ContentAnalysis/SDSSCA.html

3.2 Process of Content Analysis

This content analysis study includes four stages. In stage I, we collected SDSS-related publications from the NASA ADS (adswww.harvard.edu) digital library, which primarily archives astronomy and astrophysics documents. We retrieved SDSS-related documents up to Sept. 26th, 2009 via searching "**SDSS** OR (**Sloan** AND **Survey**)" in the documents' titles and abstracts. A total of 2,632 records of publications were retrieved. Then 200 papers, all refereed, were randomly sampled and full texts of these 200 papers were downloaded for content analysis.

In stage II, 20 papers were chosen from the 200 samples and were carefully studied to design the code book. We also conducted a small focus group discussion with three astronomers to identify the major coding contents, such as what the unit of analysis is and what should or could be coded.

In stage III, two additional coders were recruited. Both coders are senior students and majored in astronomy. The two coders were required to carefully read the code book. The authors answered questions about the code book raised by the two coders. Then in a pilot study, the first author and the two coders separately coded the 20 sample papers used in initial code book design. Inter-coder reliability of coding results was checked. The three coders then discussed the major inconsistent results and agreed on new coding rules.

In the final stage, all 200 sample papers were coded. Second round inter-coder reliability was checked between the two coders and the first author after coding 30 papers. Then a discussion was conducted to improve consistency among coders.

4 Results

4.1 Used Data or Not, and Data Sources

Table 1 shows the number of data sources used in the 200 samples. The majority (169 out of 200 papers) directly used observational data to conduct their research. Within these studies, more than half (98 papers) used one data source only, while other 71 papers used two to five data sources.

Table 1. If use of data and the number of data sources

Data sources	Num. of papers	Percent
Not use observational data	31	15.5%
Used data	169	84.5%
Used 1 data source	98	49.0%
Used 2 data sources	48	24.0%
Used 3 data sources	16	8.0%
Used 4 data sources	5	2.5%
Used 5 data sources	2	1.0%

There are total 278 data source instances used in the 169 samples, which belong to 47 unique data sources. Table 2 lists the top 10 frequently used data sources. Unsurprisingly, SDSS is the most often used data source. Authors' own data collection is the second, followed by data referred to previous studies. Two other large scale sky survey projects, 2MASS(Two Micron All Sky Survey) and 2dFGRS (2dF Galaxy Redshift Survey) were used seven and five times respectively.

Table 2. Top 10 frequently used data sources

Name of data source	Freq	Perc	Max number of objects in its database	Project completed time
SDSS	130	47.4%	300 millions	2008
Their own data collection	58	21.2%	Varied, mainly <1000	-
Referred to previous studies	26	9.5%	Varied	-
2MASS	7	2.6%	4 millions	2003
2dFGRS	5	1.8%	400,000	2003
FIRST	3	1.1%	816,331	2008
Chandra	2	0.7%	Varied	a
GALEX	2	0.7%	32180	a
RASS	2	0.7%	N*10,000	2001
UKIDSS	2	0.7%	140 millions	2005-2012

a. Based on satellite telescopes, no specific completion time.

4.2 The Number of Objects for Final Analysis

In the 169 papers that directly used observational data, we identified the number of objects that were used for final analysis. Most papers explicitly gave this number. A few (12 papers) have no description about this information. The distribution of the number of objects is skewed, with median of 102, and ranging from one to 12 million. The categorized values (see Figure 1) show that the number of analyzed objects in the 169 papers spread over this range. 72 papers (42.6%) analyzed less than 100 objects, and 10 percent papers (18 out of 169) analyzed hundreds of objects and same as those analyzed thousands of objects. There are five papers that analyzed more than one million objects.

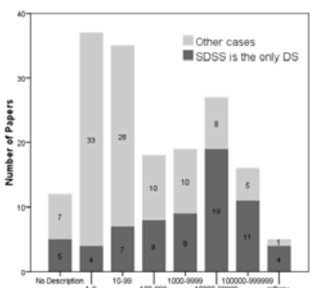


Fig. 1. Papers that use SDSS as the only data source versus other cases in the categorized number of objects analyzed

SDSS as the only data source versus other cases

Because SDSS covers millions of objects, we are particularly interested in studies that use SDSS as the only data source, and the comparison of these papers with other cases. Figure 1 shows the distribution of studies that used SDSS as the only data source versus other cases. There are 67 papers that solely relied on SDSS data out of the 169 papers that used data. It is clear that in studies using SDSS as the only data source, the majority of them used more than ten thousand objects. In other cases, studies are more like to use a small number of objects, typically less than 100 objects.

4.3 Co-usage of Data Sources

This subsection relates how the 47 data sources were co-used in the 169 papers that directly used observational data. We identified the frequently co-used data sources.

For two co-used data sources, in the 169 papers, the "SDSS" and "Their own data collection" are the most common pair-wised co-used data sources (35 times). The "SDSS" and "Referred to previous studies" is the second common pair-wised co-used pattern (15 times). The other seven data sources in the top 10 list all had been co-used with the "SDSS" with the frequencies listed in table 2. For example, "2MASS" was used seven times and in all seven times it was co-used with the "SDSS". The other six followed the same pattern.

In terms of three data sources that were co-used in one paper, the "SDSS," "Their own data collection," and "2MASS" have been co-used five times, becoming the most common three co-used data sources. The second most common three co-used group is the "SDSS", "Their own data collection", and "Referred to previous studies" (four times). In the top 10 list, only "2dFGRS", "FIRST", and "UKIDSS" have been co-used with the "SDSS" and "2MASS" each one time. The total number of studies that can use three or more large scale data sources, such as "SDSS" and "2dFGRS", are small (five papers).

5 Discussion

Several patterns associated with data sources, the number of objects, and co-used scenarios could be inferred from the above results to answer our research questions.

5.1 Data Sources

Our results demonstrate that half the SDSS studies in our sample rely on one data source only, the majority (64 out of 98 papers) relying solely on SDSS. Using multiple data sources, particular using more than two large scale data sources, is not common.

The causes of this phenomenon could be that SDSS data set alone is good enough to answer many research questions given its large coverage of sky. On the other hand, problems caused by data provenance and trust issues [16, 17] among multiple data sources could prevent researchers from using more data sources. Data from "Their own data collection" and "Referred to previous studies" normally result in few or no problems of data provenance and trust. Therefore it is common to see the two data sources appearing in studies.

5.2 The Number of Objects

Our results identify two different aspects of usage patterns of the number of objects analyzed in SDSS related publications. When considering these studies in general, the number of object is distributed from one to millions, with most papers focused on a small number of objects (<100). But when differentiating studies that solely used SDSS data from the others, the scenario is different (see Figure 1). These studies leverage the large amount of data produced by SDSS, and the majority analyzed more than ten thousand objects.

One possible explanation of this phenomenon could be the small number of cross-matched objects between different data sources. For instance, "Their own data collection" normally means follow-up observations of the selected SDSS objects by using telescopes on the ground and in space. Due to the limited observation time assigned to each telescope time application, the number of objects collected by researchers is normally small, less than 1,000. In addition, some large scale data survey, such as 2dFGRS, covered different sky areas from SDSS. As a result, the number of cross-matched objects between SDSS and other data sources is less than ten thousands.

In addition, the usability of large scale data sources could decrease the number of objects that were used. When data volume exceeds the capability of an investigator to inspect and manipulate data, which occurs at a level of several thousand objects, perhaps the biggest challenge is usability [18]. One astronomer also mentioned this issue in our focus group discussion. He mentioned that when he tried to use one large scale data source, it was hard to know how to use a large number of data efficiently and effectively when help from the data managers was unavailable.

5.3 Data Source Co-usage Models

Based on the results of the co-usage patterns and qualitative observations of paper contents, we inferred two data source co-usage models in SDSS related studies, namely a follow-up usage model and a cross-match usage model.

Follow-up data usage model

The follow-up data usage model is the most common SDSS co-usage data source model given the fact that most "Their own data collection" occurrences are follow-up observations of candidates selected from SDSS. Figure 2 depicts the data usage processes in this model. The basic stages of data usage follow the arrows. When SDSS data were released to the astronomy community, researchers queried the database for objects of interest to them, or they checked for unusual objects in the database, e.g. those marked as "unknown". Then researchers generated a set of objects as candidates for follow-up observations as they believed these candidates may lead to new discoveries. A follow-up observation was then conducted. Combining their own data with SDSS data, new discoveries were reported, or new catalogs of certain objects were created. Around 30 papers in our samples follow this model.

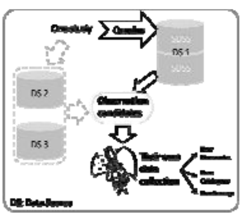


Fig. 2. Follow-up model of co-used data sources

In the follow-up model, users of SDSS data are not only consumers. They are also data producers! Their follow-up observations produce data with better quality than SDSS, data that the SDSS instruments cannot capture. However, these data normally are stored in researchers' local machines. It is difficult for researchers other than themselves to benefit from these data if no retrieval mechanisms are available. Designer and managers of data infrastructures need to consider this interesting phenomenon and provide mechanism for linking new observations of an object to its counterpart in existing data sources. A possible solution could be a Wikipedia-style knowledge center where researchers can upload their own data for certain objects and offer links to these objects in existing databases.

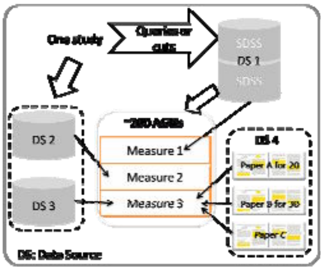


Fig. 3. Cross-match model of co-used data sources

Cross-match usage model

Because SDSS data covered only optical wavelengths, other data sources that cover other wavelengths were used to cross-match the same objects among different data sources. Figure 3 depicts the data usage process in the cross-match model. The basic process of this model is to find the different properties of the same set of objects across several data sources that focus on different properties, such as near-infrared images, x-ray images, and ultraviolet images.

The cross-match model has been strongly promoted in data-intensive astronomy [19]. Several projects such as the World Wide Telescope [20], National Virtual Observatory and International Virtual Observatory endeavor to implement the model. However, in our results papers following the cross-match model are rare, particularly for studies that cross-matched three or more data sources (five papers only). As discussed in the previous subsections, issues of data provenance, trust, and usability may prevent researchers from doing research following this model. It is apparent that much work needs to be done to enable such "virtual observatories" for science.

6 Conclusion

In the context of data-intensive science this study endeavors to model scientific data user experience. We conducted a content analysis of astronomical data usage behaviors described in SDSS publications. We conclude that:

1. Content analysis method as demonstrated in this study can model users' experiences and could be a complementary method for researchers in HCI community.
2. Although SDSS produced a large volume of astronomical data, users did not fully utilize these data. Studies that analyzed a small number of objects are the norm. This study suggests that usability issues become essential for allowing data infrastructure to be used widely in the new paradigm.
3. The follow-up usage model is the most common co-usage model in SDSS related publications. Users are not only consumers of scientific data. They are also data producers! Design of future data infrastructures need to consider this phenomenon and better support data use in this model.
4. Studies following the cross-match model are relatively rare. Issues of data provenance, trust, and usability may prevent researchers from doing this kind of research. Further HCI study of detailed usability issues associated with data infrastructures in the new paradigm is eagerly needed.

References

1. Hey, T., Tansley, S., Tolle, K. (eds.): The Fourth-Paradigm: Data-Intensive Scientific Discovery, 2nd edn. Microsoft Research, Redmond (2009)
2. Fischer, G.: User Modeling in Human-Computer Interaction. User Modeling and User-Adapted Interaction 11(65), 65–86 (2001)
3. Borgman, C.L.: Scholarship in the Digital Age: Information, Infrastructure, and the Internet. The MIT Press, Cambridge (2007)
4. Atkinson, M., De Roure, D.: Data-intensive Reseach: Making best use of research data (Draft 1), e-Science Institute (2009)
5. York, D.G., et al.: The Sloan Digital Sky Survey: Technical summary. Astronomical Journal 120, 1579–1587 (2000)
6. Bell, G., Hey, T., Szalay, A.: Beyond the Data Deluge. Science, 1297–1298 (2009)
7. Ailamaki, A., Kantere, V., Dash, D.: Managing Scientific Data. Communication of the ACM, 68–78 (2010)
8. Wallis, J.C., et al.: Digital Libraries for Scientific Data Discovery and Reuse: From Vision to Practical Reality. In: IEEE/ACM Joint Conference on Digital Library (JCDL 2010), Surfer's Paradise, Australia, pp. 333–340 (2010)
9. Van de Sompel, H., Lagoze, C.: All Aboard: Toward a Machine-Friendly Scholarly Communication System. In: Hey, A.J.G., Tansley, S., Tolle, K. (eds.) The Fourth Paradigm: Data-intensive Scientific Discovery, pp. 193–199. Microsoft, Redmong (2009)
10. Mangiafico, P.: We are all curators now: envisioning new roles for research libraries in an era of information abundance, social networks, and digital ephemera. Personal Communication (2010)
11. Doraimani, S., Iamnitchi, A.: File Grouping for Scientific Data Management: Lessons from Experimenting with Real Traces. In: The 17th IEEE International Symposium on High Performance Distributed Computing (HPDC 2008), pp. 153–164. ACM, Boston (2008)
12. Yao, Q., An, A.: Characterizing Database User's Access Patterns. In: Galindo, F., Takizawa, M., Traunmüller, R. (eds.) DEXA 2004. LNCS, vol. 3180, pp. 528–538. Springer, Heidelberg (2004)
13. McKechine, L.E.F., et al.: Research method trends in human information literature. New Review of Information Behaviour Research 3, 113–125 (2002)
14. White, M.D., Marsh, E.E.: Content Analysis: A flexible methodology. Library Trends 55(1), 22–45 (2006)
15. Lombard, M., Snyder-Duch, J., Bracken, C.: Content analysis in mass communication: Assessment and Reporting of Intercoder Reliability. Human Communication Research 28, 587–604 (2002)
16. Mann, B.: Some Data Derivation and Provenance issues in Astronomy. In: Data Provenance/Derivation Workshop, Chicago, IL (2002)
17. Balazinska, M., et al.: Data Management in the Worldwide Sensor Web. Pervasive, 30–40 (2007)
18. To share or not to share: Publication and quality assurance of reseach data outputs. Research Information Network (2008)
19. Borne, K.D., Astroinformatics: A 21st Century Approach to Astronomy, in Position Paper, George Mason University (2009)
20. Szalay, A., Gray, J.: The World-Wide Telescope. Science, 2037–2040 (2001)

Scenario and Task Based Interview to Evaluate Usability of Computer Assisted Data Collection

Luiz Agner[1], Patricia Tavares[1,2], and Simone Bacellar Leal Ferreira[2]

[1] IBGE - Instituto Brasileiro de Geografia e Estatística,
Rio de Janeiro, RJ, Brasil
[2] UNIRIO, Departamento de Informática Aplicada,
Universidade Federal do Estado do Rio de Janeiro,
Rio de Janeiro, RJ, Brasil
{luizagner,pztavares}@gmail.com, simone@uniriotec.br

Abstract. This article aims to present the method of usability evaluation called Scenario and Task Based Interviews (STBI). The method was proposed to add flexibility to field usability testing, so that they could be applied to the context of The Brazilian Institute of Geography and Statistics (IBGE). IBGE is the institute of Brazilian central Administration that performs the Census and other important official demographic and economic data collection. This evaluation technique was specifically designed to be implemented with the participation of interviewers who use PDA (personal digital assistants) to perform data collection for statistical research in Brazil. The authors analyzed the usability of the application developed for PDA to support the Continuous National Household Sample Survey (Continuous PNAD). The method proposed in this paper represented a mix of four approaches to usability evaluation.

Keywords: usability, PDA, method, data collection, interaction design, statistics.

1 Introduction

With the advancement of information technology, people and organizations increasingly need systems built with quality. Considering it is through the interfaces that people communicate with systems to perform their tasks, they must increasingly be designed with focus on usability [10].

There are different methods of usability evaluation: there are those without the presence of users - the methods of inspection or analytical methods - and there are those that involve users, called observation methods or tests with users. These can be made in the context of use or in a monitored environment such as usability labs [7].

The method proposed in this paper represents a mix of four approaches to evaluation: ethnographic observations, usability testing using a portable lab, semi-structured interviews and heuristic inspection.

In the first phase of the proposed method, the users were observed in their primary context of use: using the PDA in field to interview citizens in their homes. The second phase consisted of an observation made in a semi-controlled environment, where users

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 349–358, 2011.

were interviewed and observed in their workplace, performing tasks that simulate real situations that frequently occur in their field activities. These tasks were performed using a portable usability laboratory, where all interactions were videotaped.

The evaluation method was developed and applied for the assessment of the usability of software developed to support interviewers during the Continuous National Household Sample Survey (Continuous PNAD) conducted by The Brazilian Institute of Geography and Statistics (IBGE), the statistical institute of Brazilian central government. IBGE is the institute that performs the Census and other important official demographic and economic data collection.

The applied method, called Scenario and Task Based Interviews (STBI), sought to preserve essential characteristics of the scientific method to lend credibility to its findings and to presentation of results. It was similar (but not identical) to field usability testing presented in human-computer interaction literature [3].

This method was created to meet the conditions and peculiarities of the specific context and had been previously experienced in evaluating the usability of a 2010 Census application [2].

2 Continuous PNAD Survey

Continuous PNAD is a new survey which aims to permit a continuing research on labor and income of the population. It is the result of the merger of two polls: the Monthly Employment Survey and the National Household Sample Survey. IBGE began testing its system and methodology since October 2009 in the states of Pará, Pernambuco, Rio de Janeiro, Sao Paulo, Rio Grande do Sul and Distrito Federal, continuing in 2010, in Rio de Janeiro. In 2011, the survey will be officially launched across the country [9].

Fig. 1. Graphical interface for the Continuous PNAD mobile device: preliminary design – version 1.0.9

Continuous PNAD will be conducted from a sample of approximately 179,000 households and has several questions about work and income. It also will investigate additional subjects such as: adult and youth education, migration, vocational education, child labor, fertility, social mobility, marriage, health, food safety, information technology and communication, income and time use.

An electronic questionnaire has been created to support data collection and management. Our study examined the ease of use of the application developed for mobile device in order to generate design recommendations to make its interface more efficient and appropriate for users (Fig. 1). This study was also part of a master degree dissertation carried out in Universidade Federal do Estado do Rio de Janeiro.

3 Technology in Context of Data Collection

According to Greene [8], the introduction of a technology can transform the context of an interview to collect statistical data contributing to the feeling that this is an important event for the informant. In some cases, curiosity about the technology can draw more people to observe or participate in interviews. Sometimes it helps to "break the ice", making it easier to talk with other members of family or community. In some rare situations, technology can inspire fear or anxiety.

The methods of data collection assisted by computers are known by the terms CADC (Computer-Assisted Data Collection) or the European term CADAC, CASIC (Computer-Assisted Survey Information Collection), and CAI (Computer-Assisted Interviewing). Traditional methods of paper and pencils are often denoted by PAPI (Paper-And-Pencil Interviewing).

The main feature of the interviews supported by computer is that questions are oriented in the correct order, following an algorithm available in the interactive program on computer screen. The software has the intelligence to reconfigure the order and type of questions, based on previous answers or the information it already has on the respondent. Responses are directly entered on the computer by interviewer or interviewee [11]. As an alternative to paper questionnaires, the CADC is well accepted by interviewees and interviewers. It permits data quality improvement, especially when complex questionnaires are used. In general, respondents react positively to computer use during an interview: they attribute a greater degree of professionalism to the survey. Social interaction with the interviewer is described as comfortable [11].

4 Methodology

STBI study is basically an applied qualitative research and has six steps: (a) Literature and document research, (b) Ethnographic observation, (c) Sample selection, (d) Scenario and task based lab sessions, (e) Analysis, (f) Communication of results. These steps will be briefly described as follows:

4.1 Literature and Document Research

Initially, the team sought to understand the principles of usability with a focus on the usability of mobile devices. We studied and selected some tools and software that support professional interface evaluation, such as video recorders and applications that sync PDA with the notebook, as well as the Continuous PNAD questionnaire manuals, and videos on best practices to approach the informant.

4.2 Ethnographic Observation

Authors have scheduled a field work follow-up, where the application could be observed in real use. Our intention was to identify problems and understand the difficulties and advantages of collecting data using the PDA. To understand how the users perform their field work, the authors followed three teams of interviewers in the household interviews during the Continuous PNAD test.

Firstly, authors participated in a field visit in the Leblon neighborhood, a high middle class residential area. Then we followed a work team in various districts in Rio de Janeiro suburbs. Then the authors undertook a visit to an IBGE data collection agency in Centro to interview the employees. After that, a field visit was undertaken to Tanguá (Fig. 2), within the State of Rio de Janeiro, an area with rural characteristics, difficult transportation and no cell service.

Moreover, the authors followed a 2010 Census interviewer using a similar device in middle class area and slum of Belo Horizonte. Ethnographic observations were recorded through written notes, video and photos, presenting the view of the interviewer about his/her experience using the electronic questionnaire.

According to Cooper [6], contextual studies developed in the process of interaction design should bring the spirit of ethnographic research and apply it on a micro level.

Ethnography is not a method, is more properly a category of research in Human Computer Interaction. The ethnographic study is a powerful means of identifying the true nature of the work. It is very common for users to perform their tasks differently from what was prescribed. Ethnographic methods can discover valuable and unusual facts that would never be identified by in-house methods [4].

Fig. 2. Record of ethnographic observation: Continuous PNAD interviewer focuses on a selected household in Tanguá, a rural characteristic sector in Rio de Janeiro

4.3 Sample Selection

To analyze the target audience of Continuous PNAD workers, it was developed an online questionnaire with ten closed questions on the user's profile, their experience with technology and other surveys. It was asked whether or not he/she would like to participate in usability interviews. We chose to define our sample of participants with six individuals carefully recruited from the population of actual users. The online questionnaire was included in the Continuous PNAD management system. Users were informed about the benefits that would be obtained with usability research.

The online questionnaire was available for 30 days. After this period, results were aggregated and analyzed, yielding a total of 57 responses, where 29 users (51%) said they would like to participate. The general profile was predominantly male (70%) aged between 18 and 29 years (66%) with incomplete higher education (44%) or complete (39%), with high experience in PDA (33%) or moderately high (33%).

Based on given profile, we selected two females and four males who were coursing graduation in history, social welfare, veterinary, geography or biology. Scheduling was conducted by telephone and the interviews occurred in the Rio de Janeiro unit.

4.4 Scenario and Task Based Interview Lab Sessions

The term Scenario and Task Based Interview (STBI) was coined to avoid creating anxiety in the participants and to avoid suggesting that they were being tested. In addition, participants are accustomed to the term "interview" because of their own work in Continuous PNAD.

Comprehension of informant approach situations represented a valuable aid to build scenarios of use (common situations that occur in field work, written in the vocabulary of the user). Meetings with application development team were held to consolidate eight scenarios/tasks addressing the main field situations (Table 1). It was also provided a fictitious database to help performing the tasks.

The type of usability lab used to support lab sessions was the minimalist portable setup [13]. There is no specific room dedicated to testing: the equipment and software are taken to different locations in a notebook. Among the advantages, it is easier to recruit participants. This configuration is considered by Rubin and Chrisnell [13] as the most suitable for organizations that are beginning to test because it presents the best cost-benefit, dispensing a physical plant.

In STBI lab sessions, participants were encouraged to externalize their thoughts as they worked on tasks and activities were recorded (think-aloud protocol). As the device was synchronized with the notebook, an image of the PDA screen appeared on the screen of the notebook. A camera captured the user reactions as the whole process of usability evaluation was recorded by audio and video software.

During sessions, users were provided with the option to stop and freely register their storytelling, comments, criticisms or suggestions. So the method differs from traditional usability testing which is focused on the measurement of performance. It is essentially a method of evaluation that generates qualitative data and insights.

Table 1. Scenarios and tasks created for STBI lab sessions

Scenario	Task
You are in the field, conducting Continuous PNAD interviews. You opened a home and were informed that there lives a family with three people.	Create in the application a family formed by these three persons and confirm the information in the PDA.
By continuing to interview, you discovered that there is one extra resident to be recorded in this household.	Change the relationship of people and add grandmother Iracema with the given profile.
On the same day, you continued your work and opened another home, now with five people forming two families.	Create two families formed by these five people.
You spoke with another informant and concluded that one of persons should be excluded from the list.	Delete Octavio's brother Severino from the list of people who lives in the household.
You are now going to continue an interview that had begun in the previous month. Thus, you will need to open that file in the backup area.	Save up the current interview and open the previous month file.
You noticed that you would need to correct the information already recorded.	Change the income data of residents according to given figures.
During the same interview, you had to consult some answers to make sure they were correct.	Visualize what was the answer to the question about the number of hours worked by Mauro during reference week.
You had to stop the interview at the request of the informant. You have to call another day to complete your questionnaire.	Note the informant telephone number and type in the appropriate field a brief reminder to call him/her next Thursday between 8:00am and noon.

Upon completion of all tasks, participants answered a post-test questionnaire consisting of eleven closed and three opened questions. Based on these responses, the researchers performed an open interview, recorded on audio, giving the user the opportunity to explain their suggestions further.

In each session, two researchers were present observing and noting problems, comments and behaviors. Some sessions were attended by an invited member of application development team.

4.5 Analysis

STBI study adopted top-down analysis: six user sessions produced 48 videos which were carefully reviewed by researchers who conducted heuristic inspection to identify all problems and related suggestions by the users.

After registering all problems and suggestions, researchers used a top-down data grouping strategy beginning with Nielsen's ten heuristics categories [12] which describe general usability principles. Top-down approach began from his well-known range of established principles that could provide consistency to the analysis and to the interpretation of data [1].

5 Research Results

5.1 Ethnographic Observations Results

The authors could verify that PDA user needs to withstand harsh environmental conditions (heat, cold, humidity, drought, and light) depending on local usage. It is true that mobility imposes physical, visual and cognitive limitations to users [5]. Added to these factors, the difficult access to households (high-risk areas, distant places) and upper middle class condominiums where more and more restrictions to interviewers are imposed due to security concerns. To register these circumstances and facts the authors wrote ethnographic reports pointing out main events [14]. The color contrast of screen in sunlight is the most often cited usability problem: researchers photographed the device under the incidence of sunlight to emphasize the severity of the problem (Fig. 3).

We could also observe that ethnographic study is a powerful mean of identifying the true nature of the work. It is very common for users to perform their tasks differently from what was prescribed. In the case of computer-aided data collection ethnography can be useful to show if users spell out their questions using informal or popular language, so different from what is proposed in the official electronic questionnaire. Or, again, you will notice if users try to make a bypass of the normal operation of the application in order to avoid slowdowns, crashing or other technical issues that may impact usability.

Fig. 3. Field use observation: the device under the sunlight has legibility impaired

5.2 Portable Lab Sessions Results

Using post-task questionnaires, application was thoroughly evaluated by users. They considered it was simple, easy to learn and use. Also, 67% pointed out the simplicity, 67% rated its menus, buttons and functions as simple to use, and 50% answered that the messages are well written. Despite these figures, interactions in videos showed that some usability principles were violated to varying degrees.

Usability problems recorded in these videos were also analyzed through a heuristic inspection and were grouped together according to Nielsen´s heuristics categories [12]. Violated heuristics were identified as the following (figures show the percentage of occurrences): Match between system and the real world (13%); User control and freedom (11%); Visibility of system status (10%); Consistency and standards (8%); Flexibility and efficiency of use (6%); Aesthetic and minimalist design (6%); Error prevention (3%); Recognition rather than recall (3%); and Help users recognize, diagnose, and recover from errors (2%).

In addition to these usability principles, the analysis and sorting of results showed us that four extra categories should be created to permit a better classification of our findings. Thus, the previous list was added with 4 specific categories in order to better describe our usability issues and emphasize some important findings. See table 2 which also indicates their percentage of occurrence. It is important to note that these issues were detected during the testing phase and they were carefully considered for subsequent correction by development team.

6 Final Considerations

The impact of computer-aided interviews on data quality has been systematically evaluated by statistical institutes in diverse countries. This is an evaluation study based on Human-Computer Interaction theory. This article presented a method for usability evaluation consisting of scenario and task based interviews designed to be applied along the interviewers who used PDA (personal digital assistants) during the experimental phase of the Continuous National Household Sample Survey (Continuous PNAD).

The proposed method had essentially two major steps: observing users in their context of use (ethnographic observations) and performing common tasks while users are interviewed and interact with the interface in a notebook lab. Our experience has shown that ethnography can be considered a very important phase for the interface design because it reveals the true nature of work and avoids misunderstanding or idealization.

In our ethnographies it became evident that the users - not infrequently - conduct their work in a different way from what had been prescribed. For example, they tend to formulate questions using popular colloquial words, instead of reading the official version. They also reverse the order of questions as posed by the software, because they intend to reduce its time duration. They may also want to make it less tedious and more natural to the informant. Evidences from field studies reinforce the conclusions reached in STBI sessions in a semi-controlled environment. This allows us to affirm that both techniques have great potential when applied in a complementary manner as shown.

The method was considered quite satisfactory, since it involved usability testing of low cost. Problems were detected during the early testing phase and carefully considered for subsequent correction by development team. Moreover, it is important to note that data analysis led us to create four extra heuristic categories, where the most outstanding is concept usability - that addresses difficulties of interviewers and informants to deal with specific questions of Continuous PNAD.

Table 2. Additional heuristic categories in the Continuous PNAD - testing phase

Additional Categories	**Examples of Associated Problems**
Response Time (11%)	Slow response of the virtual keyboard.
	Slow opening of the residents frame.
	Response time beyond acceptable (30 seconds).
Explicitness, grammar and spelling (2%)	Messages in English (not in Portuguese) and containing misspellings.
	Some abbreviations, titles or text labels are incorrect or incomplete.
Bugs and crashes (19%)	System crash.
	Bug occurs when a resident is deleted.
Concept usability (6%)	Confusion between the concepts of the family chief and principal head of household.
	Interviewer and informant have difficulty to answer the question about color or race.

At the phase of reporting the results, besides scientific papers, the conclusions of this study were presented at seminars for other software development teams, aiming to spread and institutionalize usability methodology in order to contribute to make the procedures for data collection more efficient, comfortable and safe.

Acknowledgement

The authors acknowledge the support of the IBGE Executive Board, the teams of the Rio de Janeiro unit and DI-Codes, without whom this research would not have been possible. The opinions expressed in this article are the sole responsibility of the authors and do not necessarily express the views of the IBGE.

References

[1] Agner, L.: Arquitetura de Informação e Governo Eletrônico: Diálogo Cidadãos-Estado na World Wide Web – Estudo de Caso e Avaliação Ergonômica de Usabilidade de Interfaces Humano-Computador. Rio de Janeiro. Tese de Doutorado – Departamento de Artes e Design, Pontifícia Universidade Católica do Rio de Janeiro, 354p (2007)

[2] Agner, L.: A experiência do usuário no dispositivo de mão do Censo Demográfico de 2010. 9º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design. Universidade Anhembi Morumbi. São Paulo. 13 a 16 de outubro (2010)

[3] Barnum, C.M.: Usability testing and research, 428 p. Pearson Education, New York (2002)

[4] Burke, J., Kirk, A.: Choosing Human-Computer Interaction (HCI) Appropriate Research Methods – Ethnographic Methods, University of Maryland (2001), `http://otal.umd.edu/hci-rm/ethno.html`

[5] Burzagli, L., Billi, M., Palchetti, E., Catarci, T., Santucci, G., Bertini, E.: Accessibility and Usability Evaluation of MAIS Designer: A New Design Tool for Mobile Services. In: Stephanidis, C. (ed.) HCII 2007, Part II. LNCS, vol. 4555, pp. 275–284. Springer, Heidelberg (2007)
[6] Cooper, A., Reimann, R., Cronin, D.: About Face 3: The Essentials of Interaction Design. Wiley Publishing, Inc., Indianapolis (2007)
[7] Dias, C.: Usabilidade na Web: criando portais mais acessíveis, 2^a edn. Alta Books, RJ (2007)
[8] Greene, P.D.: Handheld computers as tools for writing and managing field data. Field Methods 13(2), 181–197 (2001), http://fmx.sagepub.com/content/13/2/181
[9] IBGE - Instituto Brasileiro de Geografia e Estatística. Quinto Fórum do Sistema Integrado de Pesquisas Domiciliares por Amostragem, http://www.ibge.gov.br/home/estat'istica/indicadores/sipd/quinto_forum/PNAD_Continua_Questionario.pdf
[10] Leal Ferreira, S., Nunes, R.: e-Usabilidade. LTC, Rio de Janeiro (2008)
[11] Leeuw, E.D., Hox, J.J., Snijkers, G.: The effect of computer-assisted interviewing on data quality. A review. In: Blyth, B. (ed.) Market Research and Information Technology. ESOMAR Monogaph, pp. 173–198. Esomar, Amsterdam, http://igitur-archive.library.uu.nl/dissertations/2006-1024-200101/c3.pdf
[12] Nielsen, J.: Ten usability heuristics, http://www.useit.com/papers/heuristic/heuristic_list.html
[13] Rubin, J., Chrisnell, D.: Handbook of usability testing: How to plan, design and conduct effective tests. Wiley Publishing, Inc., Indianapolis (2008)
[14] Tavares, P., Agner, L., Leal Ferreira, S.B.: Avaliação de Usabilidade de Dispositivos Móveis de Coleta de Dados Domiciliares Através de Entrevistas Baseadas em Cenários e Tarefas. In: IHC 2010 – IX Simpósio Sobre Fatores Humanos em Sistemas Computacionais, Belo Horizonte, MG, Brasil, October 5-8 (2010)

A Camera-Aided Legibility Assessment Protocol of Displays for Enhanced Human-Computer Interaction

Hongyi Cai

Civil, Environmental & Architectural Engineering, University of Kansas,
1530 W. 15th Street, Lawrence, Kansas 66045, USA
hycai@ku.edu

Abstract. Legible text and graphics presented on computer displays and projection screens are essential for benefit of human-computer interaction. Legibility of characters depends on the display brightness, luminance contrast of characters, character size, font types, color, viewing distance and angle, and observer's acuity level. In sign and display industry, a legibility index, defined as the viewing distance divided by the character height, has been widely used for legibility evaluation. However, this index fails to examine all major factors other than geometry. To enhance human-computer interaction, a quantitative legibility evaluation method, which takes into account all major affecting factors, is needed for quick and reliable guidance, goal of this study. This study thus developed a legibility assessment protocol based on a redefined legibility index as the inverse square root of the solid angle subtended by the target, a legibility equation, and innovative camera-aided high dynamic range photogrammetric techniques the author recently developed.

Keywords: human-computer interaction, camera, high dynamic range photogrammetric techniques, legibility, equation, index.

Signs

LI	Legibility index for perpendicular viewing
LI'	Redefined legibility index for both perpendicular and non-perpendicular viewing
ω	Solid angle subtended by any text and graphics to the retina of the observer's eyes
Ξ	Incident angle between the display normal and the sightline of an observer
D	Viewing distance from the observer to the display viewed at angles
H	Normal text height
V	Visual angle subtended by the height of the character
H/S_w	Height-to-strokewidth-ratio of text
S_d	Denominator in the Snellen ratio of the observer's acuity level
L_b	Background luminance (i.e., the luminance of displays)
XYZ	World coordinates
$X'Y'Z'$	Local coordinates of target plane
xz	Pixel coordinate on HDR image
κ	Camera yaw angle along Z axis
η	Camera pitch angle along X axis
φ	Camera roll angle along the shooting line (Y axis)
θ	Target yaw angle along Z axis
τ	Target vertical pitch angle along X axis
ρ	Light source roll angle along Y axis
Φ	Horizontal off-axis viewing from target to camera
α	Vertical off-axis viewing angle from target to camera

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 359–367, 2011.

1 Introduction

Legible text and graphics presented on computer displays and projection screens are essential for benefit of human-computer interaction. Legibility of characters depends on many factors, including the display brightness (i.e., the background luminance of the characters), luminance contrast of characters, character size, font types, color, viewing distance and angle, and observer's acuity level. Some empirical guidelines for legible displays are available in the literature. Table 1 summarizes these guidelines for defining an optimal cone area for locating seats with legible view of images presented on projection screens and video monitors. The horizontal fan-shaped areas were defined using a viewing angle φ, which is the angle between the display normal and the edge of the optimal view cone, and an optimum viewing distance d, which is in proportion to the display size (usually width is used) [1, 2, 3, 4, 5, 6, 7]. In Table 1, note that the size of characters is not used by the guidelines for prediction, neither lighting, nor viewers' acuity, which are all essential for legibility.

Table 1. Fan-shaped ideal viewing area defined using the existing guidelines

Display types	Sources	Fan-shaped ideal viewing area (d is viewing distance, φ is horizontal viewing angle, w is display width)	Recommendation basis
Projection screens	Aschoff [2]	2w ≤ d ≤ 6w, φ = ±30°, with maximum elevation angle of 35°, and depression angle of 12°.	DIN108
	Hauf et al. [1, 7]	2w ≤ d ≤ 6-10w, φ = ±30°-60°, with maximum elevation angle of 15°.	No experimental evidence
	Conway [8]	1.5-2w ≤ d ≤ 6w	None given
	McGowan & Kruse [9]	1.5w ≤ d ≤ 4-5w, φ = ±45°	None given
	Online projection calculator [10]	1.3w ≤ d ≤ 4.5-6w	Manufacturer supplied data
Video monitors	Hauf et al. [1]	4w ≤ d ≤ 12w (14w for less optimum condition), φ = ±35°-40° (±45° for less optimum condition), with the maximum elevation angle of 15° to the bottom of image (30° for less optimum condition).	No experimental evidence
	Allen et al. [4]	1.5w (optimum 2w) ≤ d ≤ 6w (or 4w for electronic projection), d_{max} ≤ 1 ft/in of TV diagonal size; a maximum elevation angle of 35°.	None given
	McGowan & Kruse [9]	2-2.5w ≤ d ≤ 6-7w, φ = ±45°	None given

The effect of display resolutions on the legibility of characters was also considered in the literature. The minimum acceptable displays resolution guided in 2001 was about 58 dpi [8], which is not a concern any more with today's display technologies. Sizes of characters were also considered in the literature. Human eyes can recognize the smallest characters subtended 1 minute of arc, yet minimum of 16, better 20 to 22 minutes of arc (or arcminute) were recommended [8]. For example, Table 2 listed the minimum character height recommended by ANSI standard for legibility [8].

Table 2. Minimum character height for legibility and the distance-to-height ratios [8]

Viewing distance D (feet)	Minimum character height H (inches)		D/H ratio (ft/in)	
	16 arcminute	21 arcminute	16 arcmi-nute	21 arcminute
5	0.28	0.37	17.9	13.5
10	0.56	0.73	17.9	13.7
15	0.84	1.10	17.9	13.6
20	1.12	1.47	17.9	13.6
25	1.40	1.83	17.9	13.7
30	1.68	2.20	17.9	13.6
35	1.96	2.57	17.9	13.6
40	2.24	2.93	17.9	13.7

To quantify all the efforts, a legibility index has been widely used in sign and display industry for evaluation of legibility levels of text and graphics. This legibility index, defined as the distance at which material can be read with perfect accuracy (the legibility distance) divided by the character height, as (1). This ratio equals the inverse tangent of the visual angle subtended by the character. Table 2 also shows *D/H* ratios in a practical unit of ft/in, which are commonly used in sign industry. Manufacturers of computer projectors often recommended in their product manuals optimized settings for legible screens based on such *D/H* ratios. However, this definition assumes that the viewing material is perpendicular to the observer, which is not always true. In practice, people often view computer displays and projection screens at different angles. Also, the conventional legibility index does not take into account other critical factors except for viewing distance and character height.

$$LI = D/H = 1/\tan(V) \tag{1}$$

To examine the legibility of characters viewed not perpendicular to the display, Cai and Green [9] redefined the legibility index as the inverse square root of the solid angle subtended by the target, as (2). This new definition was proven in laboratory experiments in that the three-dimensional solid angle, rather than the two-dimensional visual angle, both subtended by the character, captures how people recognize text and graphics that usually have two significant dimensions (width and height) [9]. In addition, Cai [10] proposed a new legibility equation, as (3) based on existing Howett's equation [11], which was later proven in two laboratory experiments using human subjects [10]. Equation (3) predicts legibility of text presented on assumed matte displays surfaces without glare sources visible at the peripheral of the field of view, and recognized by young observers at threshold (just readable) 100% accuracy for incident angles from 0° to 82.8°, which include nearly all possible viewing angles in human-computer interactions. Equation (3) has taken into account nearly all the major factors of legibility, including incident angle, height, distance, height-to-strokewidth ratio, Snellen visual acuity, background luminance, and luminance contrast.

$$LI' = \sqrt{1/\omega} \tag{2}$$

$$H = \begin{cases} 4.1\times10^{-4} \cdot \dfrac{H}{S_w} \cdot D \cdot S_d \cdot L_b^{-0.213} \cdot C_{\%}^{-0.532} \cdot (\cos\xi)^{-0.5} & 0^\circ \le \xi \le 65.7^\circ \\ 4.1\times10^{-4} \cdot \dfrac{H}{S_w} \cdot D \cdot S_d \cdot L_b^{-0.213} \cdot C_{\%}^{-0.532} \cdot (\cos\xi)^{-0.5} \cdot (0.024\xi - 0.577) & 65.7^\circ < \xi \le 82.8^\circ \end{cases} \tag{3}$$

2 Research Problems

To enhance human-computer interaction, a quantitative legibility evaluation method is needed for quick and reliable guidance, goal of this study. Such evaluation method will use equations (2) and (3) that the author has developed. However, for calculating legibility of displays, inputs to those two equations would be hindered by the conventional light and geometry measurement techniques using light meters, rules, etc.

The conventional way of light measurement by using meters, such as Minolta illuminance meter T-10 and luminance meter LS-100, has not dramatically changed in the past 70 years, since the first light meter, probably the AVO Smethurst High-Light exposure meter for photography, was made in 1937 [12], although meters have been continuously improved for enhanced precision and reduced size. Conventionally, light meters are either held on hands or mounted on tripod to measure often non-uniform light distributed on light-emitting and non-luminous surfaces. This is a point-by-point process, very tedious, basically incapable to measure large areas or the entire scene. In addition, those meters read an average value of illuminance or luminance at every measuring point, by simply treating non-uniform light fallen on photosensors of meters as uniform. These limitations have negative effect on lighting profession and hamper its development. Measuring common non-uniform light is thus beyond the capability of conventional meters. Basically, meters are incapable for common dynamic human-computer interaction where viewing angles and interior lighting conditions keep changing.

This situation was changed recently (2000s). Emerging high dynamic range (HDR) photography techniques could acquire luminance distribution of an entire static scene within 1-2 minutes, which is the camera shooting time, in extremely high measuring resolution [13]. Such innovative techniques use inexpensive consumer grade digital cameras, often fitted with wide-angle lens (e.g., fisheye and zoom lens), as the test rig. First, a series of low dynamic range (LDR) photographs are taken with sequential exposures, which are then fused into an HDR image by using some data-fusion software, such as Photosphere, hdrgen, Radiance, and Photolux [13]. The HDR image encloses all luminance data of the entire test scenario at pixel level. Depending on the pixel resolution of the camera sensor (e.g., 7,962,624 pixels for Canon EOS Rebel XT, 17,915,904 pixels for Canon EOS Rebel T2i), each HDR image may have acquired millions or billions of luminance values of the test scenario. Therefore, the HDR photography techniques have great potentials to overcome those limitations of conventional light measurement using meters. The HDR photography techniques have been proven reliable with acceptable accuracy in a recent study carried out by Cai and Chung [13] and Inanici [14].

Unfortunately, two inherent problems of the HDR photography techniques have hindered their applications for light measurement [15]. First, most wide-angle lenses, in particular fisheye lenses, have radical lens distortion, as shown in Fig. 1, including barrel and pin-cushion distortions. Distorted geometries of luminous elements, which cannot be corrected by using the HDR photography techniques solely, thus encumber the application of HDR images for legibility assessment. Second, except for luminance, HDR photography techniques cannot measure geometries of luminous elements required for lighting quality and legibility assessment. The coordinates in real world of any pixel on the HDR images cannot be identified by HDR photography techniques.

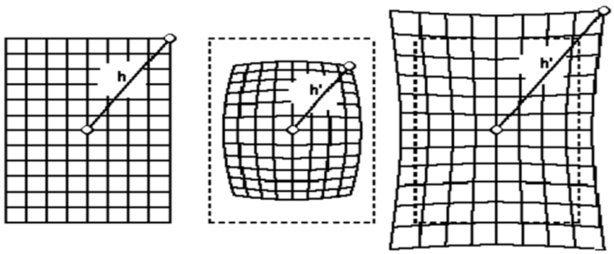


Fig. 1. Radical lens distortion [16]

3 HDR Photogrammetric Techniques

To solve this problem, this study employed an innovative camera-aided measurement method – high dynamic range (HDR) photogrammetric techniques recently developed by the author [15]. Such HDR photogrammetric techniques integrate long established photogrammetric techniques for geometric measurement into the emerging HDR photography techniques to measure both luminance and geometries of the entire scene at pixel level in a quick, reliable, and holistic way. Applied in human-computer interaction, the HDR photogrammetric techniques can measure both light and geometries simultaneously across the entire human-computer interfaces in 1-2 minutes. An HDR image fused using HDR photogrammetric techniques encloses all luminance and geometric value across the entire scene at pixel level. Such innovative HDR photogrammetric techniques will surely stimulate applications in human-computer interactions. The HDR photogrammetric coordinates and equations are introduced below.

3.1 HDR Photogrammetric Coordinates

The collinearity relationship between image and object coordinates used to develop the HDR photogrammetric techniques are illustrated in Fig. 2 [15]. As shown in Fig. 2, a consumer grade digital camera is mounted at the focal point O (0, 0, 0) (assumed at zero point (0, 0, 0), could be at any point O (X_o,Y_o, Z_o)), with yaw angle κ, pitch angle η, and roll angle φ related to the world coordinates XYZ, following the right hand rule. The HDR image plane, located on the image sensor of the camera, has two dimensional pixel coordinates xz. The principal point is located at c(x_c, z_c). The target plane is the one in blue color, with local coordinates X'Y'Z', which also has yaw angle θ, pitch angle τ, and roll angle ρ in light of the world coordinates XYZ, also following the right hand rule. The target P (X, Y, Z) and reference point P_i (X_i, Y_i, Z_i) are both located on the target plane. The position of reference point P_i (X_i, Y_i, Z_i) in world coordinates is measured in the field. Minimum three, ideally four reference points are needed for each target plane. Based on the photogrammetric coordinates, the location of target P (X, Y, Z) in world coordinates can be calculated from its pixel location p (x, z) on the HDR image, aided by the reference point P_i (X_i, Y_i, Z_i), by using some photogrammetric equations (4) – (12), to be introduced next.

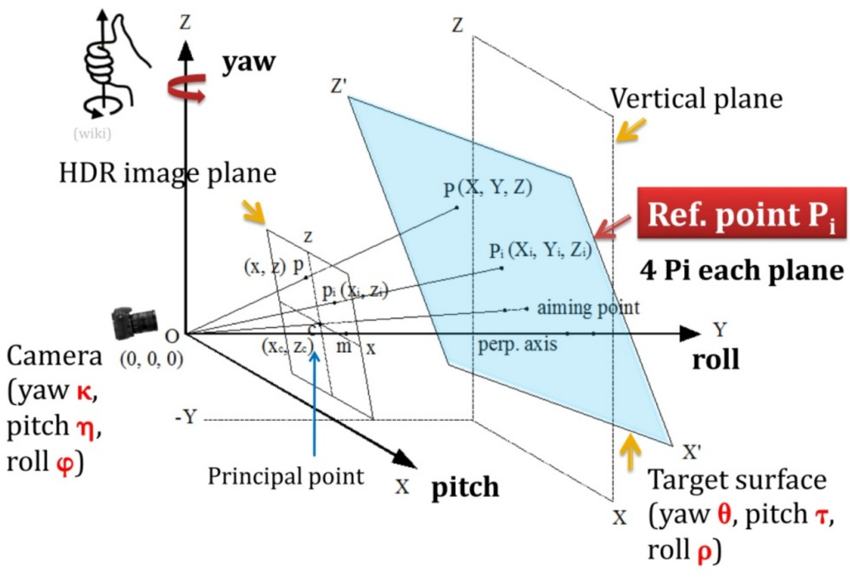


Fig. 2. HDR photogrammetric coordinates [15]

3.2 HDR Photogrammetric Equations

Step 1: find target point P (X', Y', Z') in local coordinates X'Y'Z' on the target plane

Convert target point P (X, Y, Z) in world coordinates to P (X', Y', Z') in local coordinates on the target plane [17].

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix}^{-1} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix} = \begin{pmatrix} cos\rho cos\theta & -cos\rho sin\theta & sin\rho \\ cos\tau sin\theta + sin\tau sin\rho cos\theta & cos\tau cos\theta - sin\tau sin\rho sin\theta & -sin\tau cos\rho \\ sin\tau sin\theta - cos\tau sin\rho cos\theta & sin\tau cos\theta + cos\tau sin\rho sin\theta & cos\tau cos\rho \end{pmatrix} \tag{5}$$

Since for orthogonal matrix Q, its transpose is equal to its inverse, $Q^{-1} = Q^T$, which applies in these situations [17].

$$\begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix}^{-1} = \begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix}^{T}$$
$$= \begin{pmatrix} cos\rho cos\theta & cos\tau sin\theta + sin\tau sin\rho cos\theta & sin\tau sin\theta - cos\tau sin\rho cos\theta \\ -cos\rho sin\theta & cos\tau cos\theta - sin\tau sin\rho sin\theta & sin\tau cos\theta + cos\tau sin\rho sin\theta \\ sin\rho & -sin\tau cos\rho & cos\tau cos\rho \end{pmatrix} \tag{6}$$

For reference point P_i (X_i, Y_i, Z_i), located on the target plane with yaw angle θ, pitch angle τ, and roll angle ρ, its coordinates transformation from current XYZ coordinates to X'Y'Z' coordinates is below

$$\begin{pmatrix} X_i' \\ Y_i' \\ Z_i' \end{pmatrix} = \begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix}^{T} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} \tag{7}$$

Likewise, the camera position O (X_0, Y_0, Z_0) in real world coordinates XYZ can be transformed to X'Y'Z' coordinates as below. If O (X_0, Y_0, Z_0) = (0, 0, 0), new coordinates will remain at O (X_0', Y_0', Z_0') = (0, 0, 0)

$$\begin{pmatrix} X_0' \\ Y_0' \\ Z_0' \end{pmatrix} = \begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix}^{T} \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} \tag{8}$$

Therefore, target point P (X', Y', Z') in local X'Y'Z' coordinates on the target plane is calculated as below

$$X' = X_0' + (Y' - Y_0')\frac{R_{11}(x - x_c) + R_{12}f + R_{13}(z - z_c)}{R_{21}(x - x_c) + R_{22}f + R_{23}(z - z_c)}$$

$$Z' = Z_0' + (Y' - Y_0')\frac{R_{31}(x - x_c) + R_{32}f + R_{33}(z - z_c)}{R_{21}(x - x_c) + R_{22}f + R_{23}(z - z_c)}$$

$$Y' = Y_i' \tag{9}$$

obtained from reference point P_i (X_i', Y_i', Z_i'), expressed in X'Y'Z' coordinates, where reference point is located on the target plane

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} = \begin{pmatrix} \cos(\varphi-\rho)\cos(\kappa-\theta) & -\cos(\varphi-\rho)\sin(\kappa-\theta) & \sin(\varphi-\rho) \\ \cos(\eta-\tau)\sin(\kappa-\theta)+\sin(\eta-\tau)\sin(\varphi-\rho)\cos(\kappa-\theta) & \cos(\eta-\tau)\cos(\kappa-\theta)-\sin(\eta-\tau)\sin(\varphi-\rho)\sin(\kappa-\theta) & -\sin(\eta-\tau)\cos(\varphi-\rho) \\ \sin(\eta-\tau)\sin(\kappa-\theta)-\cos(\eta-\tau)\sin(\varphi-\rho)\cos(\kappa-\theta) & \sin(\eta-\tau)\cos(\kappa-\theta)+\cos(\eta-\tau)\sin(\varphi-\rho)\sin(\kappa-\theta) & \cos(\eta-\tau)\cos(\varphi-\rho) \end{pmatrix} \quad (10)$$

Step 2: Convert P (X', Y', Z') in local coordinates X'Y'Z' on the target plane to P (X, Y, Z) in world coordinates

Finally, convert the calculated target P (X', Y', Z') in X'Y'Z' coordinates back to P (X, Y, Z) in XYZ coordinates

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix} \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} \quad (11)$$

$$\begin{pmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{pmatrix} = \begin{pmatrix} \cos\rho\cos\theta & -\cos\rho\sin\theta & \sin\rho \\ \cos\tau\sin\theta+\sin\tau\sin\rho\cos\theta & \cos\tau\cos\theta-\sin\tau\sin\rho\sin\theta & -\sin\tau\cos\rho \\ \sin\tau\sin\theta-\cos\tau\sin\rho\cos\theta & \sin\tau\cos\theta+\cos\tau\sin\rho\sin\theta & \cos\tau\cos\rho \end{pmatrix} \quad (12)$$

Using above photogrammetric equations, also aided by the reference points P_i (X_i, Y_i, Z_i) measured in the field, the real world position of target P (X, Y, Z) in the test scene can be calculated from its pixel location p (x, z) on the HDR image. At the same time, the luminance value L of pixel p (x, z) can also be obtained from the HDR image. Therefore, with known luminance and geometry, legibility of characters presented on the displays can be evaluated at pixel level.

4 Conclusions

A quick and precise legibility assessment protocol of displays aided by the two legibility equations and the HDR photogrammetric techniques is thus proposed in this study for enhanced human-computer interaction, by using a consumer grade digital camera to take some photographs of the displays. HDR images fused with the HDR photogrammetric techniques can provide both luminance and world coordinates of any pixel on the screens, for legibility assessment of displays. As expected, such innovative HDR photogrammetric techniques will have a great chance to supplement or even replace the conventional legibility assessment method using meters and rules, etc. Aided by this protocol, it is also possible to develop a quick method for post-occupancy evaluation of legible human-computer interfaces. Yet it is worthy of mentioning that further studies are necessary to improve the legibility equation (3) for specular surfaces that most displays actually have.

References

1. Hauf, H.D., Koppes, W.F., Green, A.C., Gassman, M.C.: New Space for Learning-Designing College Facilities to Utilize Instructional Aids and Media. Report of a Research Project Conducted By the School Of Architecture. Rensselaer Polytechnic Institute, Troy, NY (1961)

2. Aschoff, V.: Planning the large lecture theatre. In: Duncan, C.J. (ed.) Modern Lecture Theatres, pp. 17–25. Oriel Academic, London (1966)
3. Allen, R.L., Ault, D.K., Bowen, J.T., Clabaugh, S., Cuttill, W.J., Dewitt, B.B., et al.: Design of General-Purpose Classrooms and Lecture Halls. Pennsylvania State University, University Park (1991)
4. Hauf, H.D., Koppes, W.F., Green, A.C., Gassman, M.C., Haviland, D.S.: Revised New Spaces for Learning: the Design of College Facilities to Utilize Instructional Aids and Media. Rensselaer Polytechnic Institute, Troy (1966)
5. Conway, K.: Master Classrooms: Classroom Design with Technology in Mind. University of North Carolina at Chapel Hill, Academic Computing Publications, Inc. (1990)
6. McGowan, M., Kruse, K., (ed.): Interior Graphic Standards. John Wiley & Sons, Hoboken (2003)
7. Projector central (n.d.). Projection Calculator Pro, http://www.projectorcentral.com/projection-calculator-pro.cfm (retrieved October 30, 2009)
8. Musgrave, G.: Legibility of projected information (2001), http://www.conceptron.com/articles/pdf/legibility_of_projected_information.pdf (retrieved January 15, 2011)
9. Cai, H., Green, P.A.: Legibility Index to Examine Common Viewing Situations: A New Definition Using Solid Angle. LEUKOS: The Journal of the Illuminating Engineering Society of North America 5(4), 279–295 (2009)
10. Cai, H.: Legibility Equation for Determining Ideal Viewing Areas in Lecture Halls (dissertation). University of Michigan, Ann Arbor (2008); available from ProQuest Information and Learning, Ann Arbor, MI; AAT 3304934
11. Howett, G.L.: Size of Letters Required for Visibility As a Function of ViewingDistance and Observer Visual Acuity. National bureau of standards, Washington D.C. (1983)
12. Ollinger, J. Smethurst High-Light exposure meter, http://www.jollinger.com/photo/meters/meters/avo_smethurst.html (retrieved December 24, 2010)
13. Cai, H., Chung, T.M.: Improving the Quality of High Dynamic Range Images. Lighting Research & Technology, OnlineFirst (2010)
14. Inanici, M.N.: Evaluation of high dynamic range photography as a luminance data acquisition system. Lighting Research & Technology 38(2), 123–136 (2006)
15. Cai, H., Chung, T.M.: Introduction to high dynamic range photogrammetric techniques: beginning of a new lighting stage. In: IES Annual Conference 2010, Toronto, Canada, November 7-9 (2010)
16. Das, A., Patil, D.: Efficient Measurement and Correction techniques for Lens Distortion. EE362 Final project (2006), http://scien.stanford.edu/pages/labsite/2006/psych221/projects/06/ddpatil/ (retrieved on May 14, 2010)
17. Kraus, K.: Photogrammetry - Geometry from Images And Laser Scans, 2nd edn. Walter De Gruyter, Berlin (2007)

Measuring Drivers' Dynamic Seating Experience Using Pressure Mats

Songyi Chae, Gyouhyung Kyung*, and Kyunghyun Nam

School of Design & Human Engineering, UNIST
100 Banyeon-ri, Eonyang-eup, Ulju-gun, Ulsan, Korea
ghkyung@unist.ac.kr

Abstract. The objective of this study was to find the relationship between body-seat pressure distribution and driver comfort ratings of dynamic seating experience. A total of 38 participants performed four short-term driving sessions in a commercialized vehicle. These sessions involved two driving environments (lab vs. field-based). Body-seat interface pressure data were recorded continuously during driving, and the comfort ratings of the whole body and local body parts were measured after each session. Several body-seat pressure distribution variables were proposed to improve sitting comfort.

Keywords: Comfort, Driver Seat, Pressure Distribution.

1 Introduction

Driver seat is one of major factors which determine a driver's overall comfort and musculoskeletal health. With car customers' increasing concerns about safety and comfort, designing a proper driver seat has become a more challenging issue for car manufacturers. From an ergonomic perspective, comfort plays an important role in seat design (Juijt-Evers et al., 2003, Mehta and Tewari, 2000, Zhoa and Tang, 1994), as comfortable seats can reduce driver's back pain by alleviating vibration and road shock (Troup, 1978).

Many seat comfort measurement methods have been investigated. Objective measures include posture, the number of body movements, estimations of muscle activation and muscle fatigue by electromyography (EMG), pressure at the back rest and seat pan, model-estimations of spinal loading forces, stature loss, and foot volume changes. Subjective measures involve comfort and discomfort ratings on a whole body or local body parts.

As objective measurements, pressure variables often were used with subjective comfort ratings in some studies (e.g., de Looze et al., 2003; Andreoni et al., 2002). De Looze et al (2003) investigated comfort using both subjective evaluation and objective evaluation methods. Their study used a questionnaire for subjective evaluation and body pressure distribution variables on a massage chair as objective measures, which is based on the sitting comfort evaluation index model. Dhingra et al (2003) identified important factors for tractor seat discomfort by incorporating

* Corresponding author.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 368–375, 2011.

quantitative and qualitative approaches (i.e., seat pressure distribution and seat comfort rating). In their study, however, the exclusive use of interface pressure did not successfully predicted car seat comfort. Similarly, Lee and Ferraiuolo (1993) concluded that there was no strong correlation between subjective comfort and body pressure distribution. Gyi and Porter (1999) showed that interface pressure and driving comfort had no clear relationship. More recent studies suggested to use different types of pressure variables such as total contact area, lateral load, lateral area, average pressure ratio, peak pressure ratio variables on different body parts (Kyung et al., 2008; Na et al., 2005; Andreoni et al., 2002). For preferred seat-interface pressure values, some studies suggested guidelines based on differences in gender, body mass index, stature, and seating position (e.g., Stinson, 2003). Kolich (2003) measured occupants' preferred contour and geometry characteristics of the seats and related these to anthropometric data.

The definitions of comfort and discomfort are not yet clearly defined and still on a debate (Helander & Zang, 1997; Bishu et al., 1991). Several studies tried to conceptualize these two states (Hertzberg, 1958; Floyd and Robers, 1958) Zhang et al (1996) showed that comfort and discomfort had an orthogonal association. Kyung et al (2008) measured separate comfort and discomfort using derived versions of combinations of Borg (1990) and Corlett and Bishop (1976), and found that body-seat interface pressure had a stronger relationship with overall and comfort ratings than with discomfort ratings which is in contrast to the model of de Looze (2003). To evaluate massage chair comfort, Yang et al (2009) designed Massage Chair Comfort Questionnaire based on 4 questionnaires about sitting comfort - General Comfort Rating (Shackel et al., 1969), Body Part Discomfort Rating (Corlet & Biship, 1976), Overall Comfort Index (Kolich, 1999), and Automotive Seating Discomfort Questionnaire (Smith et al., 2006). To analyze the association of pressure data and comfort rating, different methods have been used, including statistical method (Na et al., 2005; Kyung et al., 2008; Porter et al., 2003), and artificial neural network (Kolich et al., 2004).

The objective of this study was to find the relationship between body-seat pressure distribution and driver comfort ratings of dynamic seating experience.

2 Methods

2.1 Overview of Experiment and Participants

The experiment constituted four short-term driving sessions in a commercialized vehicle, Genesis 3300 (Hyundai Motor Company, Korea). Thirty-eight participants were recruited from the local community with following criteria - Drivers who had a driver's license for more than 3 years, normal or corrected-to-normal eye vision in both eyes, and no current musculoskeletal disorders, and aged between 35 and 65. The participants were divided into three groups in accordance with their statures (Table 1). Perceived subjective comfort rating data and body-seat interface pressure data were obtained from a field-based (dynamic) and three lab-based (static) driving sessions.

Table 1. Participant characteristics

Stature group	No. of participants (male, female)	Mean (SD) stature (cm)	Mean (SD) mass (kg)
Short (<165cm[a])	12 (2, 10)	155.6 (5.0)	61.9 (9.1)
Middle	16 (13, 3)	165.3 (4.5)	65.6 (7.6)
Tall (>175cm[b])	10 (10, 0)	179.0 (3.4)	85.3 (8.5)
Total	38 (25, 13)	165.7 (9.9)	69.3 (12.5)

Source: NHANES III (1994).
a18th or lower percentiles of gender-mixed population.
b90th or higher percentiles of gender-mixed population.

2.2 Variables

In the current study, stature group and driving venue were considered as independent variables. Body-interface pressure was recorded during driving sessions and subjective comfort ratings were surveyed after each session. Pressure data were collected from bilateral buttocks and backs. A whole body comfort rating, left/right shoulders and bilateral buttocks comfort ratings were used as subjective measurement. The rating scales ranged from 0 (Not at all comfortable) to 100 (Extremely comfortable).

2.3 Experimental Protocols

Two pressure mats were placed on both seat back and seat cushion and participants were seated carefully. At the first lab-based session, the instructor adjusted the steering wheel and seat, while, at the second and third lab-based sessions, participants adjusted the steering wheel and seat themselves, except the seat back angle was fixed at 25° at the time.

2.4 Data Collection Procedures and Processing

Interface pressure data were recorded by two Tekscan (South Boston, MA, USA) pressure mats (5330 CONFORMatTM). Each pressure mat consists of 1024 (32×32) thin (1.78 mm) resistive sensors that could easily conform to the contour of the seat, and measure up to 255 mmHg (5 PSI). Each mat had an active area of 471.4 mm × 471.4 mm, and sensor pitch was 14.73 mm (0.5 sensor/cm2). Pressure data were divided into four parts (Figure 1).

2.5 Data Analysis

To analyze pressure data, 27 pressure variables were made (Table 2), which were related to average contact areas and ratio, average contact pressure and ratios, and average peak contact pressure and ratio. ANOVA and MANOVA ($\alpha = 0.05$) were used and the associations between pressure variables were evaluated by principal

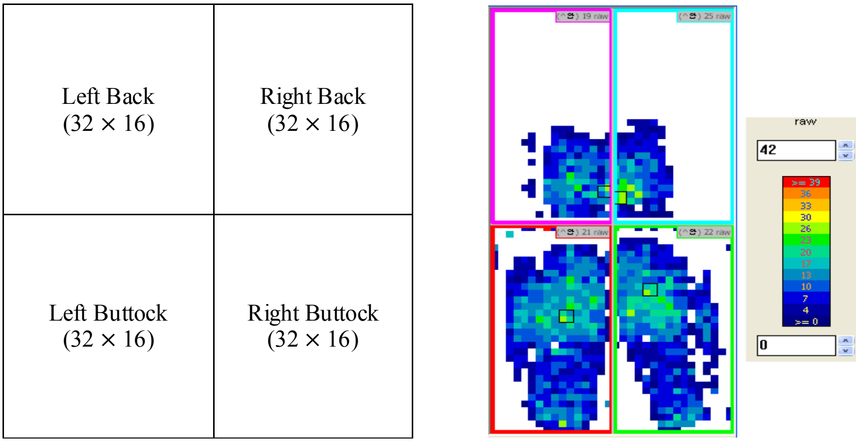


Fig. 1. Two pressure mats for four local body parts (left: number of sensors in parentheses, right: pressure distribution)

component analysis (PCA). Regression of subjective comfort rating on the factors from the PCA was used to determine a linear combination of the factors that can predict driver sitting comfort. Factors were selected by the eigenvalue (>1) and the cumulative percentage (≈90%) of variance accounted for by the selected factors (Lehman et al., 2005). The resulting factors were rotated by the varimax method.

Table 2. Pressure variables

Group	Variable name	Description
Average contact area (cm^2)	aBTL (aBTR)	Left (right) buttock
	aLB (aRB)	Left (right) back
	aSUM	aBTL + aBTR + aLB + aRB
Average contact area ratio	aBTL/aSUM, aBTR/aSUM, aLB/aSUM, aRB/aSUM	
Average contact pressure (mmHg)	avgBTL (avgBTR)	Left (right) buttock
	avgLB (avgRB)	Left (right) back
	avgSUM	avgBTL + avgBTR + avgLB + avgRB
Average contact pressure ratio	avgBTL/avgSUM, avgBTR/avgSUM, avgLB/avgSUM, avgRB/avgSUM	
Average peak contact pressure (mmHg)	pkBTL (pkBTR)	Left (right) buttock
	pkLB (pkRB)	Left (right) back
	pkSUM	pkBTL + pkBTR + pkLB + pkRB
Average peak contact pressure ratio	pkBTL/pkSUM, pkBTR/pkSUM, pkLB/pkSUM, pkRB/pkSUM	

3 Results

3.1 Effects of Stature and Driving Venue on Interface Pressures

MANOVA showed a significant ($p \leq 0.0001$) main effect of Stature and Driving Venue on the 27 pressure variables. There was no significant interaction effect of Stature × Driving Venue. From subsequent ANOVAs, stature effects were found on average contact areas, average pressure, maximum peak pressure and ratios (aRB, aLB, aBTR, aBTL, aSUM, aLB/aSUM, aBTL/aSUM, avgLB, avgLB/avgSUM, pkRB and pkBRT/pkSUM) with $p \leq 0.03$. Driving Venue effect was found only on peak contact pressure ratio (pkBTL/pkSUM). Significant mean differences between lab and field-based sessions were found for pkBTL ($p \leq 0.024$), and pkBTL/pkSUM ($p \leq 0.007$), and subject comfort rating variables for the left back and bilateral buttocks. All subject comfort ratings were higher in field-based than lab-based sessions. Significant difference of Driving Venue effects was found on the peak pressure variable at the left buttock. On lab-based sessions, the mean (SD) of the peak pressure ratio at the left buttock was 0.270 (0.084), which was lower than that on field-based venue, 0.335 (0.359).

3.2 Correlations between Subjective Comfort Ratings and Pressure Variables

Whole body and local body parts comfort ratings were higher on field-based than lab-based sessions. Comparing bilateral means of subjective comfort ratings, the left back and buttock were higher than the right. Mean average contact area was higher on the right parts. Among the whole contact area variables, right buttock ratio was the highest. As for average pressure variables, the right back and buttock were higher than the left ones as well. Right buttock ratio (avgBTR/avgSUM) was the highest. Peak pressure variables at the right back and buttock were higher than the left. pkBTR/pkSUM value was the highest one.

3.3 PCA and Regression Analysis Using Pressure Variables

With an eigenvalue > 1, nine factors were accounted for 89.0 % of the total variance (Table 3). Based on the smallest communality of 0.739 (aBTL/aSUM), the variance of each pressure variable was well accounted for by the selected factors. For Factors 1 and 4 related to contact area ratios, coefficients with opposite signs were found between the left buttock and bilateral back (i.e., aBTL/aSUM vs. a(L or R)B/aSUM), and between the back and right buttock pressure ratios (i.e., pkBTR, pkSUM vs. pk(R or L)B/pkSUM), indicating negative associations between the bilateral back and each of buttocks. For Factor 2, coefficients relating to pressure of bilateral backs and buttocks were all positive.

Regression analyses showed adjusted R2 = -0.033 between the subjective comfort ratings and the nine factors. Fitted models for comfort ratings were not significant. Based on the size of the standard beta weights (Table 4), increasing Factor 6 and Factor 7 would be effective at improving whole body comfort rating. The largest and positive beta weight for Factor 6 indicated that increasing pressure at the right buttock would be the most effective method for improving whole body comfort rating. Similarly, the second largest beta weight for Factor 7 suggested that increasing pressure at left back would be the second most effective way of improving the whole body comfort rating. Factor 4 had the third largest weights, suggesting the third strategy of increasing contact area at the buttocks.

Table 3. Nine principal components after Kaiser normalized varimax rotation (underlined values are > 0.6)

Pressure variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
	Back vs. Left buttock (area)	Back & buttock (pressure)	Back vs. Right buttock (pressure)	Buttock (area)	Back & Left buttock (pressure)	Right buttock (pressure)	Left back (pressure)	Right back (pressure)	Left buttock (pressure)
aBTR/aSUM	-.861	-.112	.163	.169	.061	.105	.129	-.073	.065
aLB	<u>.857</u>	-.002	.102	.120	.122	.157	.269	-.169	-.025
aBTL/aSUM	<u>-.764</u>	.087	.086	.103	.198	.250	.133	-.092	.038
aLB/aSUM	<u>.754</u>	-.004	.102	-.350	.141	.163	.304	-.215	-.013
aRB	<u>.701</u>	.046	.196	.336	.127	.427	-.039	.205	.077
aRB/aSUM	.561	.064	.212	-.310	.184	.528	-.052	.226	.088
avgBTR	-.022	<u>.965</u>	.005	-.003	-.130	.125	-.057	-.064	-.060
avgLB	-.077	<u>.860</u>	-.003	-.051	-.089	-.090	.371	.197	-.102
avgBTL	.139	<u>.859</u>	-.071	.100	.070	.026	.047	-.058	.235
avgRB	.063	<u>.766</u>	.021	-.119	.013	-.107	.179	.486	-.174
avgBTR/avgSUM	-.073	<u>.629</u>	.114	.038	.459	.383	-.146	-.277	.043
pkSUM	.055	.046	<u>-.854</u>	.025	-.243	.253	.120	.223	.248
pkRB/pkSUM	.034	-.025	<u>.810</u>	.099	.030	-.036	.032	.448	.233
pkLB/pkSUM	.118	.094	<u>.731</u>	.027	.030	-.124	.556	.104	.099
aBTL	-.177	.070	.032	<u>.887</u>	.115	.137	.042	-.023	-.003
aBTR	-.332	-.083	.119	<u>.876</u>	.010	.025	.045	-.007	.035
aSUM	.418	.003	-.102	<u>.838</u>	-.129	-.205	-.135	.080	-.047
avgSUM	-.067	-.117	-.010	-.105	<u>-.902</u>	.157	.038	-.054	-.041
avgBTL/avgSUM	.080	-.171	.085	.080	<u>.733</u>	.185	-.009	-.277	.382
avgRB/avgSUM	.011	-.186	.225	-.210	<u>.708</u>	.006	.182	.462	-.158
avgLB/avgSUM	-.195	.023	.171	-.105	<u>.632</u>	-.062	<u>.601</u>	.147	-.064
pkBTR/pkSUM	-.004	.058	-.248	.043	-.058	<u>.869</u>	-.096	-.130	-.083
pkBTR	.044	.023	<u>-.612</u>	.015	-.132	<u>.717</u>	.011	.017	.042
pkLB	.107	.234	-.030	.024	.006	-.066	<u>.861</u>	.262	.066
pkRB	.005	.106	.091	.072	.057	-.042	.271	<u>.865</u>	.138
pkBTL	-.075	.028	-.232	-.060	.117	-.046	.077	.073	<u>.905</u>
pkBTL/pkSUM	-.007	-.029	.496	.056	-.009	.004	-.029	.067	<u>.816</u>
Eigenvalue	3.871	3.589	2.935	2.768	2.747	2.250	1.972	1.969	1.944
Cum percent	14.336	27.630	38.501	48.753	58.926	67.258	74.564	81.855	89.056

Table 4. Regression coefficients and standard beta weights for regression models relating PCA factors to whole body comfort ratings

	Intercept	Factor1	Factor 2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9
Estimate	59.732	-1.000	1.361	1.897	-0.423	4.061	-0.566	-2.910	-1.055	-0.852
Standard beta weight	0	-0.031	0.043	2.767	2.779	2.703	2.840	2.791	-0.033	-0.027

4 Discussion

The current study focused on sitting comfort during short-term driving, and used interface pressure data as objective measures with subjective comfort rating responses. Most average pressure variables showed a difference between lab-based and field-based drivings, as field-based driving showed higher comfort rating for whole body and local body parts and comfort ratings were higher at the left back and buttocks than right parts.

Different stature groups had significantly different values of all average contact area ($p = 0.000$) and ratio ($p< 0.016$), average contact pressure in their left and right back ($p < 0.025$), average pressure ratio in their left buttock ($p = 0.016$) and peak pressure in their right back ($p = 0.000$). From a PCA, contrasting associations were found between local body parts in terms of contact area and contact pressure: negative associations between the left buttock and bilateral back from contact area and between the right buttock and bilateral back from average pressure. Driving appears to require asymmetrical body postures based on bilaterally different pressure values. The regression results suggested that increasing pressure at the right buttock would be the most effective method for improving whole body comfort rating and increasing pressure at left back would be the second most effective way of improving the whole body comfort rating.

5 Conclusions

Some pressure variables on average contact area, average and peak contact pressure ratio could be used for the assessment of driver sitting comfort. Driving experience that requires asymmetric postures could be improved by providing balanced pressure between bilateral body parts for ergonomic human-seat interface. For future study, a more comprehensive experiment is needed to include factors such as BMI, seat angles, and seating position.

References

1. Lehman, N., O'Rourke, L., Hatcher, E.J.: JMP for Basic Univariate and Multivariate Statistics: A Step-by-Step Guide. SAS Institue Inc., Cary (2005)
2. Shackel, K.D., Chidsey, P.: The assessment of chair comfort. Ergonomics 12, 269–306 (1969)

3. Metha, C.R., Tewari, V.R.: Seating discomfort for tractor operators – a critical review. International Journal of Industrial Ergonomics 25, 661–674 (2000)
4. De Looze, M.P., De Kuijt-Evers, L.F.M., Van Dieen, J.: Sitting comfort and discomfort and the relationships with objective measures. Ergonomics 46(10), 985–997 (2003)
5. Gyi, D.E., Porter, J.M.: Interface pressure and the prediction of car seat discomfort. Applied Ergonomics 30, 99–107 (1998)
6. Smith, D.R., Andrews, D.M., Wawrow, P.T.: Development and evaluation of the Automotive Seating Discomfort Questionnaire (ASDQ). International Journal of Industrial Ergonomics 36, 141–149 (2006)
7. Corlett, E.N., Bishop, R.P.: A technique for assessing postural discomfort. Ergonomics 19, 175–182 (1976)
8. Andreoni, G., Santambrogio, G.C., Rabuffetti, M., Pedotti, A.: Method for the analysis of posture and interface pressure of car drivers. Applied Ergonomics 33, 511–522 (2002)
9. Kyung, G., Nussbaum, M.A.: Driver sitting comfort and discomfort (part2): Relationships with and prediction from interface pressure. International Journal of industrial Ergonomics 38, 526–538 (2008)
10. Dhingra, H.S., Tewari, V.K., Singh, S.: Discomfort, Pressure Distribution and Safety in Operator's Seat – A Critical Review, Agricultural Engineering international: the CICG Journal of Scientific Research and Development. Invited Overview Paper 5 (2003)
11. Hertzberg, H.T.E.: Seat comfort. In: Hansen, R., Cnrnog, D.R., Hertzberg, H.T.E. (eds) Annotated Bibliography of Applied Physical Anthropology in Human Engineering, WADC Technical report 56-30, 297–300 (1958)
12. Lee, J., Ferraiuolo, P.: Seat comfort. SAE Technical Paper no. 930105 (1993)
13. Kuijt-Evers, L.F.M., Krause, F., Vink, P.: Aspects to improve cabin comfort of wheel loaders and excavators according to operators. Applied Ergonomics 34, 265–271 (2003)
14. Stinson, M.D., Armstrong, A.P., Eakin, P.: Seat-Interface Pressure: A Pilot Study of the Relationship to Gender, Body Mass Index, and Seating Position. Arch. Phys. Med. Rehabil. 84 (2003)
15. Helander, M.G., Zhang, L.: Field studies of comfort and discomfort in sitting. Ergonomics 40, 895–915 (1997)
16. Kolich, M.: Reliability and Validity of an Automobile Seat Comfort Survey, Technical Paper No. 1999-01-3232 (1999)
17. Kolich, M.: Automobile seat comfort: occupant preferences vs. anthropometric accommodation. Applied Ergonomics 34, 177–184 (2003)
18. Bishu, R.R., Hallbeck, M.S., Riley, M.W.R.: Seating comfort and its relationship to spinal profile: a pilot study. International Journal of Industrial Ergonomics 8, 89–101 (1991)
19. Na, S., Lim, S., Choi, H.S., Chung, M.K.: Evaluation of driver's discomfort and postural change using dynamic body pressure distribution. International Journal of Industrial Ergonomics 35, 1096–1985 (2005)
20. Troup, J.D.G.: Driver's back pain its prevention: A review of the postural, vibratory and muscular factors, together with the problem of transmitted road-shock. Applied Ergonomics 9(4), 207–214 (1978)
21. Floyd, W.F., Roberts, D.F.: Anatomical and physiological principles in chair and table design. Ergonomics 2, 1–16 (1958)
22. Zhang, L., Helander, M., Drury, C.: Identifying factors of comfort and discomfort. Human Factors 38(3), 377–389 (1996)
23. Zhoa, J.H., Tang, L.: An evaluation of comfort of a bus seat. Applied Ergonomics 25, 386–392 (1994)

Effects of Menu Types and Item Lengths on Operation Efficiency

Yu-Hsuan Chang and T.K. Philip Hwang

Graduate Institute of Innovation and Design, National Taipei University of Technology
1, Sec. 3, Chung-hsiao E. Rd., Taipei, 10608, Taiwan, R.O.C.
sharonchang0919@gmail.com, phwang@ntut.edu.tw

Abstract. Pop-up menus enable more efficient interface operation. Inspired from empirical inference, observation, and literature reviews, this study investigated the operation efficiency of pop-up menus through examining human's superior physical characteristics of visual search and mouse movement. A new style of menu type, Elliptic-Pie Menu, was proposed and examined on operation efficiency against traditional linear menus and (circular) pie menus while different item lengths were also analyzed. The study revealed: (1) Menu type is a significant factor of the operation efficiency; (2) Short items turn out to be more efficient than long items do; (3) Linear Menu presents the highest operation efficiency, whereas Circular-Pie Menu delivers the lowest error rate; (4) Elliptic-Pie Menu occurs significant improvement of operation efficiency with the use of short items.

Keywords: pop-up menu, pie menu, operation efficiency, menu type.

1 Introduction

As software development, more items derived from functions lead to complicated operations in searching the target item, which require great cognitive effort. Facing this challenge, pie menus [1] were proposed to save target seek time (response time) and to reduce cursor moving distance needed in contrast to traditional linear menus. Various efforts have been done to improve quality of human interface such as different menu layouts, item order, and text features. However, no previous research studied the influence of the human's physical characteristics of visual search and mouse movement as well as item lengths on pop-up menus. The goal of this study was to evaluate operation efficiency of the pop-up pie menus by taking human's physical characteristics of visual search and mouse movement into consideration. A new menu type, Elliptic-Pie Menus, was proposed and examined along with Linear Menu and Circular-Pie Menu. Eventually, operation efficiency of different menu types and effects of item lengths were delved in this study.

2 Menu Layout

Menu layout is a significant factor of menu design. Their influences on efficiency of menu operation have been proved by many studies. [1] showed that target seeking time

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 376–383, 2011.

was significant of menu operation efficiency while pie menus were inspired by Fitts' Law. Additional experimental results [2] [3] [4] also indicated that menu types had significant effects on either response time or error rate. Further, a variety of pop-up menus, for instant, slant linear menus [2], pie menus [1], and rotary menus [3], were brought out against liner menus.

3 Visual Search of the Target Item

The issue of visual search arises from the increase of mental demand imposed on users due to complicated functions.

Referring to user experiences, it is intuitive that left-to-right visual search was favored than the reverse direction. Moyle's experimental evidence [5] showed that people tended to drag mouse left- to-right more frequently regardless of selecting a target letter, word, or sentence while reading a passage.

Though researchers increasingly devoted themselves to the investigation of visual search, few research concentrated on vision movement of different item-like object arrangement. Through measurement of the eye tracking system, the results of Feng & Shen's research [6] demonstrated that people were obviously more efficient on horizontal tasks than vertical tasks. Thus, it seems reasonable that users might perform more efficiently on clicking an item target in horizontal dragging directions than that in vertical.

4 Mouse Movement and Clicking

Mouse movement is associated with human's wrist characteristics of operating the mouse. As noted by Moyle [5], vertical directional mouse movement required people to extend (up) or contract (down) their thumb and fourth or little fingers often with arm movement. Conversely, horizontal movement only involved a lateral flexing of the wrist with almost no arm movement. Furthermore, Donker & Reitsma [7] mentioned that the wrist was used to moving horizontally, so people favored move the mouse left-to-right like reading and writing. Both perspectives were supported by experimental resutls, in which Moyle's results revealed that people caused dramatically longer time and more angular errors for downward mouse movement; Donker & Reitsma pointed out people performed faster with fewer errors on horizontal dragging tasks than vertical tasks. In addition, Whisenand and Emurian [8] noted people outperformed in horizontal directions over vertical and diagonal directions.

5 Experiments

Based on previous research results, Elliptic-Pie Menu was introduced and expected to benefit by better efficiency of horizontal cursor movement and visual search. Then, experiments were conducted to examine the operation efficiency of different menu types with various item lengths.

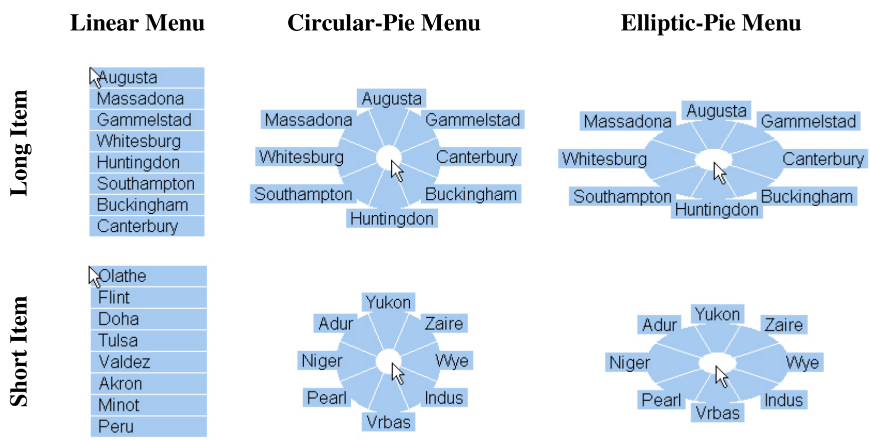
5.1 Experimental Methods

To analyze how efficient users could use the pop-up menus, performance was measured including average response time and error rate. Actually, there is a trade-off between speed and accuracy [9] [10]. Hence, the "quality of search" (Q) Score [11], originally used in the measurement of visuospatial performance, was introduced into efficiency evaluation of human interface. Hence, the influence of errors on the response time was considered to summarize the performance for different menu types.

5.2 Stimuli Design

There were six examples of menu designs contributed by three menu types and two sets of item lengths (Table 1). In each set of item length, two groups of uncommon 8 items which were referred to city names and river names were adopted to minimize the effect of familiarity. As a result, there were twelve menu layouts in experiments eventually.

Table 1. Menus Used in Experiment



5.3 Subjects and Apparatus

A total of 60 (30 females and 30 males) objects of right-hand experienced mouse users were invited. The age and educational background were not restricted.

The experiments were conducted on one Acer laptop running MS Windows XP. The screen was 14 inches with resolution of 1280x800. Three optical mice were prepared, allowing subjects to pick one based on their preferences.

5.4 Procedures

Subjects divided into 6 gender-equal groups were instructed to complete 60 tasks in a pre-determined "sequence" of menu type (for example, Linear-Circular-Elliptic) to reduce the impact of sequence effect. Thus, every menu type was executed for 20 times, including 10 long-item tasks and 10 short-item tasks. This procedure of each task was the same as regular operation without learning in advance. Subjects popped up a menu by pressing the right mouse button and then selected the given target item according to the instruction on the top of the experimental interface. During the course of operation, response time was measured in milliseconds automatically by program as well as error rate.

6 Results

36 subjects' data were selected equally from six equal-gender groups and analyzed after eliminating outliers and anomalies. Significant results among menu types were found.

6.1 Response Time Analysis

Menu type. As illustrated in Table 2, irrespective of item lengths, Linear Menu appeared to be the highest performance while Circular-Pie and Elliptic-Pie Menus presented similar operation speed on average. Results also suggested that males had advantages over females especially in operating Elliptic-Pie Menu while males and females had similar performance on the other two types of menu. A repeated measures one way ANOVA found a significant difference between menu types ($F = 7.972$, $p < 0.01$).

Table 2. Results of Mean Response Time per Task (ms) of Different Menu Types (N=36)

Menu Type		Linear	Circular-Pie	Elliptic-Pie
Mean Response Time (ms)	Female	1708.619	1895.328	1930.733
	Male	1688.392	1900.372	1869.542
	Mean	1698.506	1897.850	1900.138

Item length. As regards the factor of item lengths, on average, a short-item task required less response time than a long-item task did shown in Table 3. Males performed faster than females in both cases. The following ANOVA displayed that mean response time was found to be slightly significant between long and short items (F=3.458, p =0.065).

Table 3. Results of Mean Response Time per Task (ms) of Different Item Lengths (N=36)

Item Length		Long Item	Short Item
Mean Response Time (ms)	Female	1886.892	1802.893
	Male	1852.549	1786.323
	Mean	1869.720	1794.608

Menu type x Item length. Table 4 shows the mean response time of three menu types across two item lengths. Surprisingly, Linear Menu with short items demanded more response time than with long items, whereas Circular-Pie and Elliptic-Pie Menus favored short items. Particularly for Elliptic-Pie Menu, time was reduced dramatically by 218.172ms while an ANOVA indicated its significant difference (F=4.775, p =0.034, η^2 =0.087).

Table 4. Results of Mean Response Time per Task (ms) of Menu Types by Item Lengths (N=36)

Item Length / Menu Type	Long Item	Short Item	Mean
Linear	1672.008	1725.004	1698.506
Circular-Pie	1927.930	1867.770	1897.850
Elliptic-Pie	2009.223	1791.051	1900.138
Mean	1869.720	1794.608	

6.2 Error Rate Analysis

Menu type. The error rate analyses were proceeded for all types of menu. The results showed that Circular-Pie Menus obtained the lowest error rate of 0.97%, followed by Linear and Elliptic-Pie Menu sequentially (see Table 5). As well, males' error rate in Elliptic-Pie Menu was superior to females' while females had extremely lower error rate in Circular-Pie Menu. Referring to previous results of mean response time analyses, it would appear that males favored Elliptic-Pie Menu and achieved efficient performance.

Table 5. Results of Mean Error Rate of Different Menu Types (N=36)

Menu Type		Linear	Circular-Pie	Elliptic-Pie
Mean Error Rate	Female	0.0111	0.0056	0.0250
	Male	0.0139	0.0138	0.0167
	Mean	0.0125	0.0097	0.0208

Item length. Menus that adopted short items resulted in slightly higher error rate without difference between males and females. However, no significant result was found (see Table 6).

Table 6. Results of Mean Error Rate f Different Item Lengths (N=36)

Item Length		Long Item	Short Item
Mean Error Rate	Female	0.0101	0.0176
	Male	0.0111	0.0148
	Mean	0.0106	0.0162

In the error rate analyses, the error rate was extremely low so that its distribution did not pass the normality test, one basic assumption. Accordingly, a repeated measures one way ANOVA was not performed to examine whether menu types and item lengths had significant effects on error rate.

6.3 Efficiency Quality

However, it was manifest that Linear Menu, which had the least mean response time, did not turn out to have the lowest mean error rate. For evaluating menu operation efficiency of menu types and item lengths, *Q s*cores were administrated to examine the effects of error rate on mean response time.

$$\text{Q Score} = \frac{\text{Correct responses}}{\text{Total task number}} \times \frac{\text{Correct responses}}{\text{Total response time (sec)}}$$

Where *correct responses* are the total number of successful tasks; *total task number* is tasks executed, and *total response time* is the accumulated response time of each task, calculated in seconds (not milliseconds). Based on this formula, it is clear that higher *Q* scores reflect more efficient performance.

Menu type. Firstly, in terms of different menu types, Table 7 shows the efficiency evaluation across three menu types. Linear Menu obtained the highest *Q* score, which was consistent with the result of response time analyses. Meanwhile, Circular-Pie Menu had a slight higher *Q* score than Elliptic-Pie Menu did. Further, the significant difference between menu types was found by ANOVA ($p < 0.01$). It seems reasonable that menu type is a critical factor of influencing menu operation efficiency.

Table 7. Efficieny Evaluation of Different Menu Types

	Linear	**Circular-Pie**	**Elliptic-Pie**	***p***
Performance Score (*Q* Score)	.58	.53	.52	.001*
Correct Response (0-20)	19.75	19.81	19.58	
Total Response Time (sec)	33970.11	37957.00	38002.75	

*p <.05, **p < .001.

Item length. Secondly, the overall performances of different item lengths are displayed in Table 8. Based on *Q* scores analyses, the efficiency of short-item tasks was higher

Table 8. Efficieny Evaluation of Different Item Lengths

	Long Item	**Short Item**	***p***
Performance Score (*Q* Score)	.53	.55	.271
Correct Response	29.61(0-30)	29.53(0-30)	
Total Response Time (sec)	56091.61	53838.25	

*p <.05, **p < .001.

than that of long-item tasks. An ANOVA showed that there was no significant difference in *Q* scores between long-item and short-item tasks ($p > 0.05$). Consequently, the effects of item lengths were not as significant as the effects of menu types on menu operation.

7 Conclusion and Discussion

The experimental results showed that Linear Menu had better efficient performance according to the best *Q* score than Pie Menus (Circular-Pie and Elliptic-Pie Menus) did.

This result was supported by Samp and Decker's measurement result [12]. It presented that while the item number was fixed to six, one-level linear menus performed superior to most of various pie menus with shorter total time (response time). Moreover, Whisenand and Emurian's research [8] brought out positive response to this study indirectly. It indicated that square icon-like targets had faster movement time (response time) and lower error rate compared with circle icon-like targets.

However, the results manifestly disagreed with the research results of Jack Callahan et al. [1], the initial developers of the pie menu, which asserted that pie menus had vantage of faster operation and lower error occurrence. Under the circumstance where users have adapted to linear menus on existing consumer products, it is not odd to attribute best efficiency to users' experiences of menu operation. That is, the more familiar with a specific type of menu, the more efficient it could be.

Regarding Pie Menus, Circular-Pie Menu turned out to be less efficient but better performed in error rate, and short items were probably beneficial to improve efficiency. Comparatively, Elliptic-Pie Menu had superior efficiency improvement to other menu types by being adopted short items.

References

1. Callahan, J., Hopkins, D., Weiser, M., Shneiderman, B.: An Empirical Comparison of Pie vs. Linear Menus. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 95–197 (1988)
2. Komerska, R. & Ware, C.: A Study of Haptic Linear and Pie Menus in a 3D Fish Tank VR Environment. In: Proceeding of the 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2004), pp. 224–231 (2004)
3. Zucco, J.E., Thomas, B.H., Grimmer-Somers, K., Cockburn, A.: A Comparison of Menu Configurations and Pointing Devices for Use with Wearable Computers while Mobile and Stationary. In: International Symposium on Wearable Computers (ISWC 2009), pp. 63–70 (2009)
4. Chertoff, D., Byers, R., LaViola, J.: An Exploration of Menu Techniques using a 3D Game Input Device. In: Proceedings of the Fourth International Conference on the Foundations of Digital Games (ICFDG), pp. 256–263 (2009)
5. Moyle, M.: A Flick in the Right Direction: An Evaluation of Simple Gesture Based Controls. University of Canterbury, Christchurch (2001)
6. Feng, C., Shen, M.: Task Efficiency of Different Arrangements of Objects in an Eye-movement Based User Interface. Acta Psychologica Sinica 38(4), 515–522 (2006)

7. Donker, A., Reitsma, P.: Drag-and-Drop Errors in Young Children's Use of the Mouse. Interacting with Computers 19(2), 257–266 (2007)
8. Whisenand, T.G., Emurian, H.H.: Analysis of Cursor Movements with a Mouse. Computers in Human Behavior 15, 85–103 (1999)
9. Gramopadhye, A.K., Melloy, B.J., Gopinath, M., Budgavi, M.: An Evaluation of Economic and Performance Feedback in an Inspection Task with Explicit Economic Consequences. International Journal of Industrial Ergonomics 20(4), 327–337 (1997)
10. Shneiderman, B.: Designing the User Interface – Strategies for Effective Human-Computer Interaction. Addison-Wesley, Reading (1998)
11. Geldmacher, D.S., Hills, E.C.: Effect of Stimulus Number, Target-to-Distractor Ratio, and Motor Speed on Visual Spatial Search Quality Following Traumatic Brain Injury. Brain Injury 11, 59–66 (1997)
12. Samp, K., Decker, S.: Supporting Menu Design with Radial Layouts. In: Proceedings of the International Conference on Advanced Visual Interfaces, Rome, Italy (2010)

A Systematic Evaluation of the Communicability of Online Privacy Mechanisms with Respect to Communication Privacy Management

Periambal L. Coopamootoo and Debi Ashenden

Department of Informatics & Systems Engineering,
School of Defence & Security
Cranfield University
Shrivenham, Swindon, UK, SN6 8LA
{p.coopamootoo,d.m.ashenden}@cranfield.ac.uk

Abstract. Online privacy mechanisms have not been effective in ensuring end-users' privacy. One of the main reasons is the un-usability of these mechanisms. Although past socio-psychological studies have highlighted the need for privacy in interpersonal interactions and social relationships, approaches to designing online privacy have often not considered privacy as a communication process. In this study the principles of communication privacy management (CPM) are used within semiotic inspection to examine online privacy mechanisms. We found that privacy as a communication process breaches many of the principles of CPM. We conclude that this might explain why end-users do not interact with online privacy mechanisms effectively.

1 Introduction

In the past few years, research has shown that although individuals claim to have high privacy concerns, they do not behave privately online [1]. Although online privacy mechanisms have been designed, they have not been very effective in ensuring end-users' privacy. A key reason is the un-usability of online privacy mechanisms. It is a challenging task to ensure the usability of online privacy mechanisms since privacy is part of an individual's communication behaviour and relationship building [2; 3] and is not easily transposed into online functionality. Privacy is also person and context dependent. Although past socio-psychological studies have highlighted the need for privacy in interpersonal interactions and social relationships [4; 5], approaches to designing online privacy have not considered privacy as being an interpersonal communication process.

A study was performed with the aim of determining whether privacy mechanisms are designed in a way that mirror how individuals interact in an offline environment and hence whether they offer the functionality that end-users require. We analysed the online privacy mechanisms using a widely studied and applied theory of privacy communication - Communication Privacy Management (CPM) theory [6].

CPM is an evidenced-based theory of how individuals manage both their disclosures and those granted to them by others. CPM has been used in previous research to

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 384–393, 2011.

understand how people decide to disclose private information in offline settings and also to understand and address the tension between disclosure and privacy within the ecommerce context [7]. CPM offers a practical framework which can be used to assess online systems and consists of assumptions and interaction principles [8].

The analysis for this study was performed using the semiotic inspection method (SIM) which is a theory-based analytical usability evaluation method that concentrates on the communication aspects of HCI [9]. The purpose of theory-based evaluation methods in HCI is to assess the quality of interfaces and interactions in the light of a given perspective of HCI. The primary purpose of SIM is to evaluate the communicability of interactive computer artefacts by focusing on user interface meanings expressed through design. SIM helps to examine the diversity of signs presented to end-users while they interact with HCI artefacts and places the concept of the sign at the heart of semiotic inspection [10]. The signs present in computer interfaces include words, colours, dialogue structures and graphic layouts.

In this paper we describe the usability evaluation study undertaken to explore the communicability, and therefore the usability, of three cases of privacy mechanisms. Communicability is an important aspect of usability since not only can it contribute to the understanding required for learnability, memorability, ease of use, efficiency and satisfaction but also to the understanding of the privacy features which is an important part of exercising privacy online.

2 Method

In this section, the different types of privacy designs are first introduced, followed by a brief rationale behind the selection of particular examples that will be used for the communicability evaluation. The CPM principles are then explained.

2.1 Privacy Designs

Privacy designs can be categorised by the approach used to provide privacy, for instance, the ‘notice and choice’ type provides a privacy notice, which is usually a long sequence of text that acts as liability shield for the service provider, and choice for the end-user to decide whether to be private or to disclose [11]. Another type of privacy design is permission management which can be both embedded within the system’s interaction design or added-on. A third type of privacy design is privacy protection (for example, through pseudonymity) where a trusted third-party provides end-users with a pseudonym that can be used to interact with the service provider so as not to reveal their full identity. In this study we selected an example of each of these privacy designs.

For the ‘notice and choice’ type, the Amazon web site was selected due to its wide use for e-commerce. For the privacy management type, the Internet Explorer 8 browser was selected since it is the most widely used internet browser. For the pseudonym type it was difficult to find a publicly available example. Finally, the Microsoft Dreamspark portal was selected for this type of privacy approach since it allows registered students access to services without the need to know who the students are. Moreover, verification of student status can be performed through the use of a student identity number or the use of Athens and Shibboleth institution logins. The Shibboleth scenario was used for the study. It is a standard based, web single sign-on package that can be used across, or within, organizational boundaries. It allows sites to make informed authorization decisions for individual access of protected online resources in a privacy-preserving manner [12].

2.2 Communication Privacy Management Theory

CPM theory was devised to understand the way people manage private information from a communication perspective targeting the privacy regulation that takes place during interpersonal interactions. It can be used to untangle the dialectical tension of both disclosing and maintaining privacy. In this study the theory was applied to the chosen designs to help understand and evaluate their usability. CPM theory is based on five principles of private information management that represent the organising principles interlinking both individuals and groups [13].

Principle One of CPM stipulates that disclosure and privacy form a dialectical tension; that is, if there is no disclosure, privacy mechanisms means are not required. When individuals disclose and they know they are disclosing, they usually communicate via some privacy rules.

Within Principle Two of CPM, individuals believe they own their personal information and hence have a perception of rightful ownership and a right to control it. They further believe that they are entitled to disclose information or keep it private depending on what seems the best choice for them. This leads to Principle Three of CPM, where it is stipulated that since individuals own their information and have a right to control it, they should have the means to regulate the flow of information that they define as private through the formation of privacy rules.

According to Principle Four, once private information has been disclosed, parties become responsible for co-owning and co-managing the information. This means that for a viable relationship to exist, and to allow smooth interaction between the parties involved in the disclosure, the rules must be coordinated through linkages. Linkages can be of different types depending on the balance of information shared between the parties and the balance of power that results.

Principle Five of CPM stipulates that dissimilar expectations and misunderstandings of privacy parameters can cause conflict in the handling of private information. This turbulence has to be dealt with by re-coordinating boundary rules.

It was decided that the use of SIM, bounded by CPM principles, would enable an understanding and evaluation of privacy mechanisms by teasing out the communication aspects of privacy within a systematic and rigorous analysis of the design. It has been argued that the interpretive results of SIM are objective, can be validated, and are comparable to other accepted methods because of the preparation and validation steps of the inspection method [9].

2.3 Data Collection and Analysis

SIM is systematically carried out in five steps. The first three steps consist of the inspection of documentation and ‘Help’ content together with the static interface signs and dynamic interaction signs. The questions asked at each of these steps refer to the principles of the CPM with the aim of understanding whether online privacy has been designed and communicated in a way that end-users are used to offline. The ‘Help’ content, static interface signs and dynamic interaction signs were then collated separately into three meta-communication templates.

For Amazon the ‘Help’ instances consisted of the privacy notice, the privacy paragraph in the ‘Customer promise’ and the help on the ‘Your profile’ section of the web site. The static interfaces selected were the ‘Create your profile – your page on Amazon.co.uk’ and the ‘View basket’ interface before proceeding to checkout. The pages ‘Create a profile’, ‘Place an order’ and ‘Write a review’ were chosen for the dynamic interactions.

For Internet Explorer 8, the ‘Help’ instances included ‘Change your privacy settings’, ‘Customise your privacy settings for an individual website’, ‘Understanding privacy polices’ and ‘Understanding security and privacy features’. The static interface was the only interface presented by the privacy tab whereas dynamic interactions emerging from this interface included: selection of a setting for internet zone, customisation of privacy settings for an individual website (sites feature), selection of the advance option, selection of default and checking pop-up blocker followed by settings.

For Microsoft Dreamspark, the ‘Help’ documentation selected was the privacy notice whereas for the static interfaces, the initial home page (and the home page after login in with Windows Live Id) was selected. The dynamic interaction examined consisted of the sign-in feature which is followed by verification and by download since this sequence of actions is necessary to obtain access to services and since different parties are involved that impacts how personal details are treated.

In the fourth step of SIM the templates were compared to identify consistency and contradictions, whether gaps in one template were filled by another template and whether the signs mutually reinforced each other. In the last stage the findings from the comparison were analysed with respect to the CPM principles and a substantiated judgement was made of communicative problems that may prevent end-users from understanding that they are disclosing, that privacy means are available and how they can be used and hence from effectively using these online privacy mechanisms.

3 Results

In this section, an analysis is provided for each of the cases with respect to the CPM principles. We start with the ‘notice and choice’ scenario of Amazon, followed by IE 8 privacy management and end with Microsoft Dreamspark and pseudonymity.

3.1 Amazon

As outlined in section 2.1 above Amazon provides a ‘notice and choice’ privacy design. This means that notice is provided to customers in a privacy notice and end-users have a choice to disclose or to maintain their privacy according to their needs. Data was collected for each of the sub-questions for each CPM principle and the output is shown in the table below.

Table 1. Amazon inspection

CPM Principles		Amazon		
		Help	Static	Dynamic
Principle 1	What	✓	✓ x	✓ x x
	Who	✓	✓ x	x
	How	✓	✓ x	✓ x x
Principle 2	Choice to release or not	✓ x	x	x
	Warning, notification	✓	✓ x	✓ x x
Principle3	Means	x	x	x
	Awareness of means	x	x	x
	Motivation	x	x	x
	Context	✓	✓	✓
	Risk/benefit	x	x	x
	Rule Acquire/social	✓ x	x	x
Principle 4	Who is linked	✓	✓ x	✓
	Type of links	role and susceptibility	coercive and identity	coercive
	Awareness of type	✓ but not for profile	✓ x	x
	Ownership	manipulative and obligatory	obligatory with amazon	manipulative
	Permeability	x	x	x
Principle 5	Awareness of turbulence	x	x	x
	Feedback after disclosure	x	x	x
	Means to deal with turbulence	x	x	x

From the table we can see that for Principle One, Amazon allows the end-user to be aware, and to understand, that there will be disclosure of personal information. However, the static interface provides for awareness at the time of profile creation only and even then the recipient of disclosure is general. For the dynamic interaction sequences, the end-user would know that he or she would disclose personal information when creating a profile and have a general idea to whom but will not know this at a later stage when placing an order or writing a review.

For Principle Two, the above table shows that the 'Help' templates provide information that can reveal the release of control but not of ownership. 'Help' provides information about a change in boundary through either a warning or notification whereas the static interface and dynamic interactions do so only at the point of profile creation. For profile creation across the three meta-communication templates, the end-user might know about the release of control in the 'Help' sections as there is a sentence that says, 'you can also set who gets to view it: everyone or just you' whereas in the static interface and dynamic interactions it says, 'it's up to you with whom you share what'. The 'share' word suggests control rather than release. This distinction shows an inconsistency rather than mutual reinforcement across the three meta-communication templates. Furthermore, the 'Help' link is not very evident since it is on the top right corner of the screen and the 'Need help' button is similarly far from where the user interaction happens.

For Principle Three, it can be seen that overall Amazon does not provide the means, in terms of privacy features or mechanisms, to allow end-users to regulate the flow of their disclosure, although it does provide features of information management. The only indication of a privacy feature is within profile editing where the profile can be set to be completely private or open to everyone. However, the interactions do not lead to the editing feature. Amazon, by providing a 'notice and choice' architecture, is consistent in not providing a means to regulate flow of disclosure apart from profile creation where the user is warned with 'it's up to you what you disclose to whom' but still there is no means of regulating disclosure by what is shared to whom. The only solution would be for the end-user to not give any information rather than giving away everything. Moreover, the created profile is automatically set to public, and end-user would not know that this is the case; unless he or she edits the profile. Even so the 'edit' button is on the far right of the screen and is not embedded in what is happening during profile creation.

For Principle Four, although the end-user might know who is linked within the new boundary, the types of linkages formed differ between the 'Help' function, the static interfaces and the dynamic interactions. The types of linkages in the 'Help' sections are role or susceptibility linkages whereas the static and dynamic templates are of a more coercive nature due to the lack of information that would enable understanding. Ownership of shared information through the three templates is either obligatory or manipulative in nature. Moreover, no permeability is negotiated. The lack of boundary coordination and negotiation causes end-users to unknowingly acquire Amazon rules within both the static interface and dynamic interactions. Also, since the types of links are not clear, the end-user would not really know what is shared with whom, hence cancelling out the information provided in support of Principle One.

Amazon is consistent throughout its templates in not providing for turbulence awareness and a means to deal with turbulence according to Principle Five of CPM.

3.2 Internet Explorer 8

Internet Explorer 8 as mentioned in section 2.1 is the most widely used internet browser. It provides privacy management features through its privacy tab by enabling control of access to browsing activity information that can be used to profile end-users. The table below shows the data collected for the SIM analysis for each of the questions of CPM.

Table 2 shows that IE 8 does not communicate Principle One of CPM - that is awareness of disclosure - effectively to end-users. The 'Help' section clearly provides information about what is shared but is limited to a few examples whereas the static and dynamic templates are confusing with regards to what is shared to whom and how.

Table 2 also shows that IE 8 communicates Principle Two of CPM only through the 'Help' section and documentation. IE 8 does not effectively communicate awareness of the release of control and ownership which would have happened through a choice of control/ownership versus none and also does not communicate the awareness of changing boundaries which would happen through notifications or warnings.

In the 'Help' section, information about the means to regulate disclosure is available, but there is no information about risks versus benefits. However, the static interfaces and dynamic interactions provide no information that the features are a way of regulating disclosure. The static interfaces and dynamic interactions are also consistent in not providing motivation. The absence of motivation was detected by a lack of persuasive techniques compared to the ‘Help’ sections which provide for some motivation via suggestion, reduction and tunnelling techniques. They also provide no relation to the context of disclosure; do not enable a risk/benefit assessment and do not provide granular options that would favour formation of one’s own rule combination.

Table 2. Internet Explorer 8 Inspection

CPM Principles		IE 8		
		Help	Static	Dynamic
Principle 1	What	✓	x	x
	Who	x	x	x
	How	x	x	x
Principle 2	Choice to release or not	✓	x	x
	Warning, notification	✓	x	x
Principle3	Means	✓	✓	✓
	Awareness of means	✓	x	x
	Motivation	✓	x	x
	Context	✓	x	x
	Risk/benefit	x	x	x
	Rule Acquire/social			
Principle 4	Who is linked	✓	✓	✓
	Type of links	Role if user understands		
	Awareness of type	x	x	x
	Ownership	x	x	x
	Permeability	x	x	x
Principle 5	Awareness of turbulence	x	x	x
	Feedback after disclosure	x	x	x
	Means to deal with turbulence	x	x	x

For Principle Four, the templates mention allowing or blocking cookies from ‘first’, ‘third party’ and ‘browsing’ websites. If end-users understand what cookies and these types of websites are, they might form role linkages; that is knowingly disclosing information to those with the role of providing a service and maintaining the disclosed information. However, the lack of explanation might result in end-users not understanding and so inadvertently agreeing to coercive linkages. There are also no ownership rights or permeability negotiations.

3.3 Microsoft Dreamspark

As explained in section 2.1, Microsoft Dreamspark portal was selected for the pseudonym type of privacy approach since it allows registered students access to its services through their pseudo-identity, that is, either through their registered student identity number or through a verified code from the student’s academic institution. The verification scenario used for this study is Shibboleth authorisation.

Table 3. Microsoft Dreamspark Inspection

CPM Principles		MS Dreamspark		
		Help	Static	Dynamic
Principle 1	What	✓	x	x
	Who	✓	x	x
	How	✓	x	x
Principle 2	Choice to release or not	x	x	x
	Warning, notification	x	x	x
Principle3	Means	x	x	x
	Awareness of means	x	x	x
	Motivation	x	x	x
	Context	Access to MS software		
	Risk/benefit	x	x	x
	Rule Acquire/social	x	x	x
Principle 4	Who is linked	MS	MS	MS & Institution
	Type of links	Role		
	Awareness of type	✓	x	x
	Ownership	x	x	x
	Permeability	x	x	x
Principle 5	Awareness of turbulence	x	x	x
	Feedback after disclosure	x	x	x
	Means to deal with turbulence	x	x	x

From Table 3 above, it can be seen that end-users will only understand that they are disclosing from the privacy statement and not through the static interface and the dynamic interactions. End-users will also only be aware of the type of links made from the privacy statement. The static interface and dynamic interaction sequence provides no other information about disclosure, privacy, the means to be private, boundary linkages, negotiation or turbulence. This analysis shows that this type of privacy-preserving approach controls personal details for the end-user but the end-user is not fully aware of who is involved in processing the information or how they can link personal details. For instance, it is unclear how much access Microsoft Dreamspark has to the end-user's institutional details. According to the technical aspects of Shibboleth, Microsoft would not have access but the end-user does not know that. Moreover, the amount of access that educational institutions have on the activities performed via Microsoft Dreamspark is not evident either.

4 Discussion

When the meta-communication templates from the 'Help' pages and privacy notice, the static interfaces and dynamic interactions are compared, the inconsistencies indicate that the release of meta-communication is incomplete by design, and such incompleteness contributes to end-users not understanding that they are disclosing to Amazon, IE 8 and Microsoft Dreamspark. If end-users cannot understand what they are disclosing, they cannot be expected to use privacy mechanisms and in Microsoft Dreamspark they are driven into using the mechanisms without being made aware of the fact.

Moreover, while Amazon provides some indication of disclosure (Principle One) in its profile creation feature, the means to regulate disclosure (Principle Three) is not available. Hence the end-user does not have a choice. Furthermore, while Amazon indicates changing boundaries with a warning (Principle 2) for profile creation, it does not provide a choice of being private or disclosing. This means that although end-users would be aware that other customers will have access to their data, they would not realise that they are releasing control and ownership of that data.

Both Amazon and IE 8 provide quite blurred boundary linkage hints and do not allow for ownership and permeability negotiation. End-users might not even be aware how ownership of their data is shared - that is how much control the recipient has to act on his or her data. Turbulence awareness and resolution is also completely absent.

Amazon has a ‘notice and choice’ architecture that relies on its privacy notice to provide most of its privacy information to end-users but clear links to the ‘Help’ sections are not provided in the static interfaces and dynamic interactions. Furthermore the privacy notice could make use of a layered notice using icons for instance to enhance understanding. Amazon should also allow end-users to understand the risk and benefits of their actions.

Although IE 8 provides the means to regulate disclosure in the ‘Help’ section, it provides no awareness that these are privacy means or gives the motivation to use them or explains their relation to the context of disclosure in the static interfaces and dynamic interactions. The ‘Help’ section is not directly linked to the static interfaces and dynamic interactions. The consequence is that although the means are available, end-users would not know for what and how they can be used.

Finally, although Microsoft Dreamspark, through the use of Shibboleth, provides privacy to end-users by segregating institutional authorisation and software access, privacy communication is not performed by the end-user. The approach is easy to use, simple and usable, but it has taken control away from the end-users who will not necessarily understand who has access to their data and how. In this case the usability of the privacy mechanism has been reduced to functional usability through a technological approach. While this approach works in this scenario by keeping the end-user away from any decisions, it might not work for other scenarios. The current scenario also relies heavily on the privacy statement to inform end-users about the privacy of their details.

5 Conclusion

This study provides an evaluation of three different types of privacy approaches in terms of their interaction design and highlights the flaws in the way privacy awareness and control aspects are communicated to end-users. While a solution to making online privacy mechanisms more usable and effective is complex, this study helps to identify interface and interaction points that require enhancements. The designs that were analysed have not considered disclosure and privacy as a dialectical tension; the focus on disclosure and privacy are in different parts of the designs and they either give details of disclosure without providing the means to protect privacy or they provide privacy mechanisms without facilitating an awareness of what they offer. Moreover, the designs do not make end-users aware of how ownership rights are shared let alone assist them to decide those rights for themselves. Finally, end-users are only aware of permeability rights in certain instances but cannot negotiate these. This study then provides evidence that end-users cannot be expected to interact with these online privacy mechanisms effectively as the mechanisms offered do not map to the end-users' real-world expectations. The study forms part of an ongoing research on the usability of online privacy mechanisms.

References

1. Acquisti, A., Grossklags, J.: Losses, Gains, Hyperbolic discounting: an experimental approach to information security attitudes and behaviour. In: 2nd Annual Workshop on Economics and Information Security - WEIS 2003, May 29-30. University of Maryland (2003)
2. Gerstein, R.: Intimacy and Privacy. Ethics 89, 76–81 (1978)
3. Rachels, R.: Why Privacy is important? Philosophy of Public Affairs 4, 323–333 (1975)
4. Westin, A.: Privacy and Freedom. Athenum (1967)
5. Altman, I.: The Environment and Social Behaviour: Privacy, Personal Space, Territory and Crowding. Brooks/Cole Publishing, Monterey (1975)
6. Petronio, S.: Boundaries of privacy: dialectics of disclosure. State University of New York Press, Albany (2002)
7. Metzger, M.: Communication Privacy Management in Electronic Commerce. Journal of Computer-Mediated Communication 12(2) (June 2010), http://jcmc.indiana.edu/vol12/issue2/metzger.html
8. Petronio, S., Durham, W.T.: Communication Privacy Management Theory. In: Baxtor, L.A., Braithwaite, D.O. (eds.) Engaging Theories in Interpersonal Communication: Multiple Perspectives, pp. 309–322. Sage Publications, Inc., USA (2008)
9. de Souza, C.S., Leitao, C.F., Prates, R.O., Bim, S.A., da Silva, E.J.: Can inspection methods generate valid new knowledge in HCI? The case of semiotic inspection. International Journal of Human-Computer Studies 68, 22–40 (2010)
10. de Souza, C.S., Leitao, C.F., Prates, R.O., da Silva, E.J.: The Semiotic Inspection Method. In: VII Brazilian Symposium on Human Factors in Computing Systems, Natal, RN, Brazil, November 19-22, vol. 323. ACM, New York (2006)
11. Spiekermann, S., Cranor, L.: Engineering Privacy. IEEE Transactions on Software Engineering 35(1) (January 2009)
12. Morgan, R.L., Cantor, S., Carmody, S., Hoehn, W., Klingenstein, K.: Federated Security: The Shibboleth Approach. EDUCAUSE Quarterly 27(4), 31 (2011), http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/FederatedSecurityTheShibboleth/157315
13. Petronio, S.: Communication Privacy Management: What do we know about Family Privacy Regulation? Journal of Family Theory and Review 2(3), 175–195 (2010)

User Evaluation of Internet Kiosks in University Setting

Erkan Er[1] and Kürşat Çağıltay[2]

[1] Information Systems, Middle East Technical University,
06531 Ankara, Turkey
erkaner@ii.metu.edu.tr
[2] Computer Education and Instructional Technology,
Middle East Technical University, 06531 Ankara, Turkey
kursat@metu.edu.tr

Abstract. METU kiosks are established at common points of METU campus to meet the immediate internet needs of students however they are not used at expected level. In this study, usability of METU kiosks is evaluated to identify design problems that may discourage users from using kiosks. For the evaluation, series of user trials were conducted based on some common tasks. Evaluation results show that there are critical usability problems with the design of input devices mounted on kiosk. Users generally had problems while typing with keyboard, while moving pointer using trackball mouse and while clicking using touch screen. In addition, no significant difference is observed between inexperienced and experienced users in terms of their overall success during trials. Specific examples for these usability problems and related recommendations are provided in this paper.

Keywords: Usability, user evaluation, kiosks, human computer interaction.

1 Introduction

Kiosks are generally described as centers which enables public access to information and communications technology to address needs of the community (e.g. educational, economic, or entertainment needs) [1]. Kiosk is not a new technology and its design has taken many forms for years depending on target user, its services and new technology trends. Back to its initial forms, according to Frances and Jennifer [2], kiosks that are designed in 1990s were like uninteresting boxes with simple interfaces produced for specific purposes. However, the role and nature of the kiosks has changed significantly throughout the years and today there is variety of kiosk types designed for numerous purposes. For example, there are internet kiosks for public internet access; there are photo kiosks to take photo or to print pictures; there are telekiosk used for communication. The other kiosk types can be listed as audio kiosks, credit card kiosks, customer service kiosks, educational kiosks, hotel kiosks, medical kiosks and museum kiosks. In the following table, 21st century kiosks are compared with the early kiosks in terms of dialog design, philosophy and connectivity.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 394–403, 2011.

Table 1. Comparison of early and recent kiosks [2]

	Early Kiosks	**Recent Kiosks**
Dialog Design	- Menu based access to a limited number of screens. - Touch screen.	- Web/Windows-like interfaces, with data entry dialog boxes, dropdown lists, scroll bars, pointers and hyperlinks. Touch screen supplemented by keyboard.
Philosophy	- Task based.	- Customer service based.
Connectivity	- Stand alone or connected to one proprietary database.	- Internet enabled for real time information provision and communication.

An important observation from this comparison is the change in the philosophy: there is a shift from task focus to customer focus in kiosk design. Furthermore, the interface design of kiosks has changed significantly. Whereas initial form of kiosks interfaces has offered limited number of screens consisting of limited number of navigational elements, today's kiosks supports web/windows-like complex interfaces containing dropdown lists, scroll bars, pointers and hyperlinks. The transition to more complex interfaces has been driven by mainly internet technology and web pages containing detailed information [2].

Today, the number of terminals, or kiosks, placed in public locations to deliver information and services to the general public is increasing day-by-day [3]. In parallel to this, as a public access system, usability has become an important aspect for kiosks. According to Rowley and Slack (1999), there are four major components to be considered while designing public access systems: user characteristics, environment, task and technology, and kiosk design should support the task, the user profile and the environment in which the task is to be performed. Therefore, kiosks should be designed in a way that user should complete a specific task (e.g. sending e-mail) in a desired duration using the available technology provided by the kiosk in the environment.

Recently, kiosk technology has been established in Middle East Technical University (METU) at common locations such as Student Affairs, Computer Center, Medical Center and Cafeteria to meet the immediate internet needs of students. However, usage rate is below what was expected and many of the students are even not aware of that technology. The aim of this study is to evaluate usability of these kiosks by testing whether specific internet tasks could be completed by target users using the available technology on the kiosk. For this purpose, series of user trials have been conducted for set of frequent tasks and the results have been analyzed and discussed to emphasize design problems that impair the usability.

The reminder of the paper is organized as follows: first, related literature is briefly described to identify trends in evaluation methods of kiosk usability; then methodology employed in empirical study is discussed; then the results will be presented with data analysis and findings are explained; then findings are discussed, and finally paper ends with the conclusion section.

2 Literature Review

There are various case studies in the literature related to measurement of kiosks' usability. In this section, these studies are briefly explained by focusing on especially their methodology sections to identify how they evaluate the usability of kiosk.

One of these studies is "User evaluation of the MASK kiosk" conducted by Lamel et al [5] In this study, Multimodal Multimedia Service Kiosk has been evaluated on different tasks. Task details for each user trial are recorded in terms of number of inputs, transaction time and success rate for quantitative analysis. In addition, after user trials, a questionnaire consisting of general questions about the subject and overall satisfaction has been applied. Therefore, quantitative results are supported by the qualitative data.

The other important usability study related to kiosks is the case study conducted by Kules et al [6] in 2004 to evaluate the usability of community photo library. In this study, "usage logs, a survey and three days of extensive informal observations were employed to evaluate the use of the system by an estimated 1000 users" [6]. Based on the results of the study, some design guidelines are developed, under four categories: immediate attraction, immediate learning, immediate engagement, and immediate disengagement.

Ashford et al have conducted a study, called "Electronic Public Service Delivery through Online Kiosks: The User's Perspective" [4]. In this study, mainly user's views towards using kiosks and users' awareness of kiosks have been identified with the help of questionnaire surveys with 1068 respondents. There is no user trial conducted to get quantitative data. However, usage statistics are used to obtain comparative statistics on usage levels and length of usage times.

There is another study conducted by Yi-Shun Wang and Ying-Wei Shih [7], which applies Unified Theory of Acceptance and Use of Technology by conducting survey over 244 respondents to investigate the determinants of usage behavior of information kiosks. In addition, "moderating effects of age and gender differences on the relationships between the determinants and behavioral intention/use behavior" are identified [7].

3 Methodology

In order to evaluate the usability of kiosk some user trials have been performed onsite in the Student Affairs of Middle East Technical University (METU) with the target user group, which are students of METU. Students around were invited to attend trials as potential respondents and ones those are willing to participate were directly selected as experimental without any restriction on gender, age or prior experience with METU kiosks provided that they are computer literate. Gender and experience range of subjects are shown in Table 2.

Prior to user trials, a pilot study has been conducted to identify possible problems or limitations with the initial design of the tasks for user testing. According to observations during pilot study, these tasks for user testing were reviewed and final version (See Table 3) was decided. There are seven tasks in total designed to address different usability aspects of kiosk, which are basically usability of keyboard, mouse and touch screen.

Table 2. Comparison of early and recent kiosks

Gender		**Kiosk Experience**	
Male	Female	Experienced	Inexperienced
9	6	6	9

Table 3. Details of tasks for user testing

	Task Description
Task 1	Login to system.
Task 2	Check your MetuMail.
Task 3	Send a short mail to yourself.
Task 4	Search for a book in library system.
Task 5	Learn the transcript fee for student copy.
Task 6	Repeat Task 2 with touch screen to read the mail you have just sent.
Task 7	Logout from the system.

Sub-steps for each task are defined with all details to constrain each user to follow the same procedure to accomplish the specific task. For example, user can access MetuMail, in task 2, by either typing direct URL or by clicking on MetuMail link on METU home page. For such cases, in order to prevent uncontrolled diversity of actions within tasks, users are forced to apply the specified procedure to accomplish tasks. This restriction is must to obtain consistent results in user testing.

Before testing, subjects were given a brief introduction to the purpose of the study briefly and to the tasks to be performed. Then, a small questionnaire (See Appendix A) was applied to learn demographic information as well as prior knowledge of users. During user trials, the durations for completing tasks and the number of errors made have been measured and recorded for each user and for each task separately. After trials were completed, a short interview has been conducted with users to identify their attitudes towards the use of METU kiosks and its design. These interviews provided opportunity to learn users' suggestions for alternative design for the kiosks.

4 Results

User trials were performed with 15 subjects. The main results obtained in user trials are provided in Table 4 in terms of duration and errors. In the duration column, firstly average duration times are calculated considering the timings of all subjects for each task separately. Also, percentages of average durations are calculated to identify the tasks taking longer time or shorter time in general. The errors column comprises total number of errors, percentage of errors and average number of errors per user. These values help to identify tasks with higher or lower percentage of errors and to detect tasks that challenged users most or least.

Table 4. Results of user trials for each task testing

TASK #	Duration (s)		Errors	
	Avg. Dur.	*%*	*Total # of errors*	*Avg. per user*
1	16.2	9.62	10 (7.8 %)	0.67
2	24.2	14.38	11 (8.6 %)	0.73
3	44.8	26.65	28 (22 %)	1.87
4	25.0	14.89	11 (8.6 %)	0.73
5	19.7	11.72	2 (1.5 %)	0.13
6	29.2	17.39	60 (47 %)	4.00
7	9	5.35	5 (3.9 %)	0.33

Based on the observations during user testing, a prediction is made regarding the time periods that users spent on using different input devices on the kiosk. For that purpose, a score over ten is assigned for representing the length of time spent using keyboard, trackball mouse or touch screen separately for each task. These scores are provided in the Table 5.

Table 5. Scorings of input devices for each task based on duration of usage

Tasks	Keyboard	Trackball Mouse	Touch Screen
Task 1	9	1	0
Task 2	5	5	0
Task 3	8	2	0
Task 4	7	3	0
Task 5	0	10	0
Task 6	3	0	7
Task 7	0	10	0
Total	**32**	**31**	**7**

According to scores presented in Table 5, keyboard and trackball mouse are used equally in terms of durations. On the other hand, the least used input device is the touch screen because subjects were allowed to use touch screen device only in task 6. In addition, whereas task 1, 3 and 4 are completed mainly using keyboard, task 5 and task 7 are completed solely using trackball mouse.

According to Table 4, the task that is completed in longest period of time is the task 3 with the average value of 44.87 seconds, during which users tried to send a short mail themselves. In addition, task 3 is one of the tasks that users made lots of errors to complete; 22 percentage of errors in overall occurred during the completion of task 3. This indicates that the length of the completion duration for task 3 is not related only with the nature of task but also related with the difficulty that users has with using input devices, and consequently, with the high number of errors made.

Another important observation about results is the high rate of errors in task 6. Almost half of the total errors (i.e. 47.2 %) occurred during the completion of task 6. It is surprising that although it has the highest number of errors with 60, its duration is not the highest. For instance, although its average error number per user is higher than the twice of errors occurred during task 3 (i.e. 4.00 vs. 1.87), its average completion duration is much less than the completion duration of task 3. This is actually because different input devices were used in task 3 and task 6: during task 3, users used mainly keyboard or trackball mouse whereas they used only touch screen during task 6. Since users were able to recover from their mistakes more easily by using touch screen, they could complete task 6 in shorter time despite of higher number of errors.

Furthermore, the tasks with minimum completion time are task 7, task 1 and task 5 with the values of 9 seconds, 16.2 seconds and 19.72 seconds, respectively. These tasks also were completed with the minimum error rate, 3.9 %, 7.8 % and 1.5 % respectively. Interestingly, two of these tasks (i.e. task 5 and task 7) were completed only by using trackball mouse. This means that users made less number of errors when they used trackball mouse to complete tasks. On the other hand, task 6 and task 3, which does not require mouse use, are the tasks that users made high number of errors. Task 6 is completed using mainly touch screen and task 3 is completed using mainly keyboard. Therefore, users faced some problems while using these input devices.

Table 6. User-specific results of trials

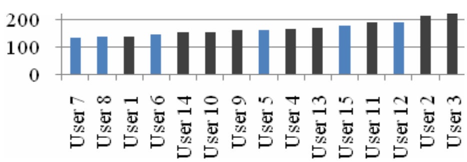
Users	**Duration**		**Errors**		**Experience**
	Total dur. (s)	*Task with max. dur.*	*# of errors*	*Task with max. error rate*	
1	138	3	12	6	No
2	214	3	12	3	No
3	248	2, 6	17	2	No
4	167	1	6	1, 6	No
5	160	6	11	6	Yes
6	145	3	4	6	Yes
7	131	3	7	6	Yes
8	136	3	11	3	Yes
9	159	4	5	4, 6	No
10	154	2	4	6	No
11	188	3	12	3	No
12	189	6	6	6	Yes
13	168	3	8	6	No
14	152	3	6	6	No
15	176	3	6	6	Yes

In Table 6, overall duration and error values are summarized for each user. In the duration column, total duration of all tasks for each user and the task that took longest time are shown. In the error column, total number of errors made by each user during all experiment is provided and also the task which is completed with the highest error rate is indicated. Beside duration and error column, experience column is included to differentiate users based on their experience with METU kiosks. “Yes” means that the user has an initial experience with METU kiosks, whereas “No” means that experiment is the first time that user used the kiosk.

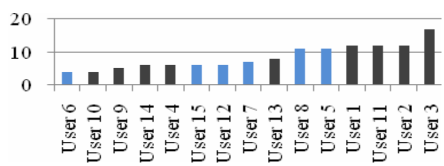
According to Table 6, nine users has completed task 3 with the longest duration during which users have used intensely keyboard. An interesting behavior of users while completing task 3, which is not provided in result tables, is that eleven users preferred to use “Tab” key on the keyboard to place cursor on the text area to type e-mail although general tendency in desktop or laptop computers is using mouse click to place cursor on specific text field. This is an indication of that users generally got easily tired while they were using trackball mouse and preferred to use keyboard as alternative when possible.

Moreover, considering error rates, task 6 is the common task during which users made errors most: eleven users made errors mostly in task 6 whether they are experienced or in experienced. Task 5 and task 7 during which users used only trackball mouse is not available in the column of "task with maximum error". This is actually another representation of the same observation in Table 3: total number of errors in task 5 and task 7 (i.e. 7) is less than the number of errors made in any other single task. Thus, users could be able to use trackball mouse free of error in general.

For further analysis, the data dispersed in Table 6 are represented in bar chart format in following figure. In the first chart (a), users are sorted ascending according to durations that they completed the experiment. In the second cart (2) users are arranged based on the number of errors they made in total in ascending order. In these figures, users with prior experience are indicated with blue color and users without experience are marked in dark green color.



(a)



(b)

Fig. 1. (a) Experiment duration for each user; (b) Total number of errors of each user

One of the significant observations from the figure is that users with prior experience could not succeed much better than the users without prior experience. Figure 1 shows that, although user 7 and user 8 completed the experiment in the shortest duration, many inexperienced users could complete the experiment in shorter duration than other experienced users. In addition, a similar case is observed in the figure 2. That is, error rates change independent from the experience level of users. For example many inexperienced users (e.g. user 10, user 9 or user 14) have made less number of errors than experienced users (e.g. user 15, user 12 or user 7) in general. Thus, there is no consistency within experienced or inexperienced users regarding duration of experiment and error rate.

Average of duration and error rate for experienced and inexperienced users, on the other hand, shows significant difference and it helps to differentiate these two user groups. These average values are presented in table 7.

Table 7. Average results for experienced and inexperienced users

User Groups	Avg. Dur. (s)	Avg. Error Rate
Exper. Users	156.17	7.50
Inexper. Users	176.44	9.11

Table 7 shows that experienced users could complete experiment in shorter time with lower error rate in overall. To be more specific, there are 20 seconds difference considering duration and almost 2 errors difference considering average error rate.

5 Discussion

In this study, user-based evaluation of kiosks is conducted based on user trials in natural setting. The purpose of the study is to test usability of input devices on the kiosk, which are keyboard, trackball mouse and touch screen. Results obtained in user trials provided enough evidences to identify problematic usability issues with the kiosk design. In addition, interviews with users have provided some clues to identify usability problems and propose related recommendations for alternative designs.

5.1 Usability Issues Related with Keyboard Design

Task 1 involved mainly use of keyboard to enter username and password information to be able to login the system and also little use of mouse to click on login button. The unusual keypad design of kiosk has challenged user in identifying intended key. Especially, almost half of the users could not enter password correctly at first time because it contains generally special characters. As an exceptional case, one of the users could not continue experiment after first task because he could not enter the password correctly. His password, which is provided by Computer Center of Metu, contains '#' character which is not available in keyboard of kiosk, which was installed by Computer Center of Metu. This means that Computer Center prohibits some of Metu students to use kiosks unintentionally.

Furthermore, keyboard design has lead to similar problems in Task 3 and Task 4, which are completed in longer duration than tasks completed on PC. In task 3, keyboard was used to login the mail account, which is a similar activity with task 1. On the other hand, task 4, which is about sending a short mail, required more extensive use of keyboard. Users made a common error while typing the receiver e-mail address: they could not find the '@' key for seconds because it is positioned in different part of kiosk keyboard, which is much different than the standard location of '@' key. Furthermore, users made many mistakes while typing the mail content and they could not easily corrected their typing mistakes.

In interview sessions, many of the users complained about the position of the keyboard: it is much lower than the level of the screen which makes eye transition between keyboard and screen difficult, and it is positioned in deep side of the kiosk which makes it harder to type. Besides its unfamiliar keypad design, these issues also have increased the time it takes to complete tasks requiring extensive use of keyboard and lead to significant time difference for the same tasks between kiosk and PC.

5.2 Usability Issues Related with Trackball Mouse Design

Trackball mouse was used mainly in task 5 and task 7 to click on buttons or links to navigate to related pages or menu items. Although results (See Table 3) show that users made less number of errors during these tasks, Table 7 indicates that these tasks took longer time than others if compared with the duration of corresponding tasks completed using PC. This is actually because users had to make many mouse movements to position pointer at intended location.

The common mistake related with mouse use is that many users could not find out where the left click button of mouse is at first. If they could not find it after initial attempts, they were informed about the place of the left click button of mouse.

Interview results show that users generally felt inconvenient with the use of trackball mouse, and they tried to use keyboard instead of mouse if possible. An observation during user trials supporting this tendency is that users generally preferred to use tab key instead of mouse to access the intended textbox field to type e-mail.

5.3 Usability Issues Related with Touch Screen

Users were not introduced to touch screen capability of kiosk until task 6 to observe whether they notice that capability without any external notification. None of the users could predict the touch screen functionality, which means that kiosk design could not enhance the visibility of the touch screen capability.

As stated before, touch screen was used only in task 6 which has the highest error rate in overall. There are two reasons for high error rate: (1) adjustment of the touch screen is not well done, which causes users not to be able to click intended part of the screen, (2) the standard view of web pages seems to be small and therefore inappropriate for touch screen use.

Efficiency of touch screen can be evaluated better if completion duration of task 6 and task 2, which comprised very similar activities, are compared. As a reminder, task 2 involved use of mouse and keyboard. Task 6 took longer time in general than task 2 although the reverse was expected because of touch screen use. This comparison again pointed to inefficiency of touch screen design.

According to interview results, although users liked the touch screen functionality they think that it could be more effective to use it. In addition, some users suggested that the monitor should not be vertically positioned; there should be some horizontality, which will enhance the usability of touch screen. To enhance functionality of touch screen, some special software can be used in kiosks instead of operating systems used in PCs. This software can provide a user interface with appropriate size navigational elements that users are able to click with higher efficiency. In addition, this software can have some intelligence so that users do not have to login to mail account after the first login to system

6 Conclusion

In this study, usability of METU kiosks was tested with the serious of user trials which comprises pre-determined tasks. The data tracked during user trials have been analyzed to elicit evidences to evaluate usability of different input devices mounted on the kiosk. Evaluation results have revealed that the design of the kiosk comprising all input devices has some usability problems and that there is no significant difference in success of experienced and inexperienced users.

There are some limitations of that study. Firstly, the number of participants for user testing is a bit low. Conducting a user testing with higher number of users such as 100 users can produce more valuable and substantial results for evaluation, which can be considered as future work. Furthermore, in that study, timing of use of each input device is estimated for each task only based on observations during user trials. However, exact time for each input device could be tracked to identify exact duration of each device use. As a final limitation, completion duration of tasks in PC are measured based on 5 computer literate users, which are different than subjects of the experiment. However, subject could perform tasks in kiosk and then in PC for better evaluation.

In summary, this study provided useful conclusions on the design problems of kiosk. These conclusions and recommendations can be considered as a usability guideline for the future design of kiosks at METU.

References

1. Kumar, R., Best, M.: Social impact and diffusion of telecenter use: A study from the Sustainable Access in Rural India project. Journal of Community Informatics 2(3) (2006)
2. Ashfort, R., Slack, F., Rowley, J.: Kiosks 21: a new role for information kiosks? International Journal of Information Management 22(1) (2001)
3. Maguire, M.C.: A review of user-interface design guidelines for public information kiosk systems. International Journal of Human Computer Studies 50(3), 253–286 (1999)
4. Ashford, R., Rowley, J., Slack, F.: Electronic Public Service Delivery through Online Kiosks: The User's Perspective. In: Traunmüller, R., Lenk, K. (eds.) EGOV 2002. LNCS, vol. 2456, pp. 169–172. Springer, Heidelberg (2002)
5. Lamel, L., Bennacef, S., Gauvain, J.L., Dartigues, H., Temem, J.N.: User Evaluation of the MASK Kiosk. Speech Communication 38, 131–139 (2000)
6. Kules, B., Kang, H., Plaisant, K., Rose, A., Shneiderman, B.: Immediate usability: a case study of public access design for a community photo library. Interacting with Computers 16(6), 1171–1193 (2004)
7. Wang, Y.-S., Shih, Y.-W.: Why Do People Use Information Kiosks? A Validation of the Unified Theory of Acceptance and Use of Technology. Government Information Quarterly 26(1), 158–165 (2009)

Evaluating Ubiquitous Media Usability Challenges: Content Transfer and Channel Switching Delays

Alexandre Fleury, Jakob Schou Pedersen, and Lars Bo Larsen

Aalborg University, Department of Electronic Systems
DK-9220, Aalborg Ø, Denmark
{amf,jsp,lbl}@es.aau.dk

Abstract. As ubiquitous media is developing rapidly, new HCI challenges emerge. In this paper, we address usability issues related to the transfer of content between fixed and mobile devices, as well as channel switching delays on mobile devices. We first provide an extensive review of the field. We then evaluate four relatively novel approaches for initiating a transfer of video content from a mobile phone to a TV screen. Seen from a user's point of view, familiarity and comfort are found to be important decision factors when selecting a preference among the proposed methods. Furthermore, we identify a threshold level above which people appear to be annoyed when switching between TV channels on a mobile device, and investigate factors that may influence the perceived acceptability of such delay.

Keywords: Mobile media, content transfer, channel switching delay, user studies, simulated environment, WoZ.

1 Introduction

TV broadcasters are no longer only focusing on traditional TV sets when broadcasting content. For instance the British Broadcasting Corporation (BBC) [1] and Zweites Deutches Fernsehen (ZDF) [2] explicitly guarantee that their audience can get free access to services in ways and on devices that suit them - acknowledging that users want potential access to media at all times. In order to facilitate this, BBC as well as ZDF have launched in 2007 large media portals: the 'BBC iPlayer' [3] and the 'ZDF Mediathek' [2] both enable users to watch live TV as well as programs from the past week from e.g. internet-connected computers, set-top boxes, and mobile phones.

Broadcasters' intention of supporting multiple platforms and devices makes good sense as ubiquitous computing is becoming increasingly widespread and popular. A 2-month study of 11 mobile information workers in a large IT company in Finland showed that work-related tasks on a daily basis were heavily distributed between a wide selection of devices (e.g. desktop PCs, laptops and various handhelds). By doing so, the workers reported to benefit in terms of efficiency, multitasking, personal ergonomics, privacy and security. [4]

Similarly, Dearman and Pierce interviewed 27 workers and found that ubiquitous computing is not only present at work but also in private homes. In average,

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 404–413, 2011.

interviewees had one laptop/desktop PC at work/school, one laptop/desktop PC at home, one cellular device and at least one portable device (typically a digital camera or an iPod). The majority of them even had a laptop PC dedicated for bringing between work/school and home on a regular basis. The interviewees argued that several reasons exist for using multiple devices: form factor, device affordances, portability, and task completion time. In addition to switching between devices for different tasks, participants reported to increasingly engage in activities that span devices (e.g. using a laptop PC in combination with a desktop PC). [5]

In both studies, synchronizing information across devices was reported as a challenge for users. In order to cope with this, they used a combination of portable media, Emailing, shared directories and server-based services.

Combining the ubiquitous computing scenarios with the broadcasting of content to different platforms and devices enables ‘ubiquitous media’ as defined in [6]. In this paper, the typical ubiquitous media environment includes a TV, a laptop PC and a Smartphone on which users can either watch on-demand- or live-TV. As reported in [4] and [5] an unresolved usability issue exists however when trying to merge a media experience across devices: the synchronizing of information.

In addition, although a lot of focus has been placed on usability of mobile TV, an open issue exists still for what concerns acceptable response times for channel switching [7]. Providing low response times comparable to those known from standard TV is important, but the definition of ‘low’ remains unclear.

In this paper we therefore address those two unresolved usability issues related to the successful integration of TV and mobile devices: Video transfer across devices and TV channel switching delays on mobile phones.

1.1 Outline

In the next section, we provide an overview of previous research within the integration of fixed and mobile devices as well as TV channel switching delays. We then describe our study design, methodology and results for our two conducted experiments addressing these issues. We finally discuss the results and applied methodologies and provide a general conclusion opening for potential future work.

2 Previous Work

2.1 Integration of Media-Displaying Devices

Ubiquitous media and multiple-device environments have recently come to the close attention of scholars of various areas. In particular, the use of various devices in the home environment has been the focus of the ethnographic study reported in [8], which investigated media habits at home. After having identified the current and ideal home media use of 27 families, the authors designed an experimental mobile device acting as a second screen to control the TV channels as well as displaying photos on the TV screen. Sharing media content and especially broadcast multimedia files (including long sequences) across devices with peers seems in fact to be one of the strongest drivers for mobile multimedia usage [9].

More concrete empirical studies have dealt with the integration of mobile devices with television sets. For instance it has been proposed to "put the EPG onto everyone's mobile phone" in order to personalize a shared TV in a typical family home [10]. According to the families interviewed during the ethnographic study, being able to access the EPG on their own mobile device and to personalize it allows family members not only to manipulate it without disturbing other TV viewers, but also to help resolving some conflicts with regard to the control of the TV.

A tendency emerging from the literature is that so far most studies have integrated mobile devices into the TV experience from a control perspective, the typical usage being the manipulation of content displayed on the TV screen from the mobile device, and the access to functionalities on the phone that complements the TV experience.

Most recently, the extensive work by Cesar and his colleagues on the concept of secondary screen illustrates the diversity of possibilities offered by multi-device media environments [11]. To this purpose, the authors have developed a taxonomy describing user behaviors in such environment. The taxonomy includes content control (what and how to consume TV content), content authoring (manipulating the TV content) and content sharing (communicating with others). Relevant to the present study, this taxonomy includes "presentation continuity", which allows users to bring their media content along on their mobile device when leaving the room in which resides the fixed TV set. According to Cesar et al., this feature has been mainly investigated through a technological perspective, disregarding user studies.

2.2 Channel Switching Delay

It is commonly accepted that channel switching delay is a critical usability issue with mobile television. The time and cognitive resources allocated to watching television on the move are limited, and users do not like to wait for neither the service to load on their mobile device nor for the channel to switch when requested [12].

This issue has been tackled in the research literature mostly from a technical perspective so far. Clues concerning user's opinion on the topic are provided by studies not specifically targeted at the user experience with channel switching delays on mobile devices. Additionally, there seems to be very little consistence between general recommendations available, technical solutions proposed and user studies performed on existing mobile TV systems, as summarized in Table 1.

Even though the last study mentioned in Table 1 focuses on IPTV and not mobile television, Kooij et al. have conducted the closest study to the one reported in this paper. The authors followed the ITU recommendation concerning the estimation of end-to-end performance in IP networks formulated in [19], and conducted a comprehensive user study validating a model that links channel zapping time and perceived quality expressed as a Mean Opinion Score (MOS). The "zapping experiment" involved 21 test subjects who rated video clips (10 seconds, no audio, video resolution of 720×575) on a web-based interface displayed on a computer screen. When switching between the clips, the test subjects experienced delays of 0, 0.1, 0.2, 0.5, 1, 2 and

5 seconds. The results from the experiment indicate a threshold of 0.43 seconds as acceptable channel switching delay. [18]

Table 1. Maximum acceptable channel switching delay according to various studies

Source	**Study type**	**Max. delay**
Nielsen, 1994 [13]	Recommendation	< 1 second (interactive systems)
ITU, 2007 [14]	Recommendation	< 2 seconds (mobile TV)
Rezaei et al., 2007 [15]	Technical	0.9 to 1.6 seconds (DVB-H)
Hsu and Hefeeda, 2009 [16]	Technical	500 ms (DVB-H burst broadcasting only)
Knoche and McCarthy, 2005 [17]	User	5 to 15 seconds (SMDB)
Cui et al., 2007 [12]	User	Up to 10 seconds (SDMB)
Koij et al., 2006 [18]	User	0.43 second (IPTV)

For what concerns the potential factors that can influence the acceptability of channel switching time, Godana et al. investigated the effect of displaying random advertisement pictures during channel switching delay ranging between 0 and 5 seconds on an IPTV system [20]. According to this subjective experiment, displaying advertisement improves the reported Quality of Experience (QoE) for transition time longer than 0.65 seconds. However, showing advertisement only postpones the threshold at which users get annoyed. For short zapping times, the authors argue that a black screen generates better QoE.

In another experiment, De Watcher et al. proposed to display a low quality version of the channel to be displayed when switching channel on a fixed digital television [21]. The authors argue that not only the perceived effect of changing channel is reduced for the user, but the method also optimizes the transition delay itself. In fact, a technical evaluation of the approach showed that it was possible to reduce the channel switching delay from 1400ms to 78ms.

With this previous work in mind, the next section presents our approach in addressing the first of the two unresolved usability issues related to the successful integration of TV and mobile devices, namely the transfer of video content from a mobile phone to a TV set.

3 Acceptability of Transfer Methods from a Mobile Phone to a TV

It seems that despite the number of technical solutions investigated to enable presentation continuity in ubiquitous computing environments, no user studies have been conducted so far to validate the approach against potential end users.

In comparison, our approach tackles the problem from the users' perspective only, regardless of technical requirements or limitations. In this purpose, our contribution is twofold: we verify the interest in transferring video content from a mobile phone to a TV, and seek to identify the preferred method from a usability perspective.

3.1 Transfer Methods

Four pre-selected sets of actions for handing over content were evaluated in terms of usability. The proposed methods were all inspired by common interaction paradigms.

Tossing. This action encompasses a method with which the user literally "tosses" the content from a mobile device to a fixed one, conceptually similar to interacting with the Nintendo Wii. Previous research has shown that 'tossing' as means of interaction is fun to use, although a bit difficult to grasp [22].

Proximity. Here the user has to physically approach a fixed device with the mobile device in order to transfer the content. Previous research has shown that users in general are willing to use 'touching' as means of interaction with devices when such devices are nearby, when security issues exist or when ambiguity is a concern [23].

Browsing. Here the user actively searches for equipment capable of taking over the presentation of content from a mobile device. Once located, the user selects a device and the handover is initiated. Previous research has shown that browsing may be seen as a very technical way of interacting with devices and that users therefore tend to avoid it when possible, unless the device in question is outside touching or pointing range [23].

Pointing. With this action (inspired by Point-and-Connect [24]), the user simply points at the device that is to take over the playback of the video from the mobile phone of the user. When pointed to a compatible device, its name appears on the mobile for the user to click on in order to initiate the transfer of content.

3.2 Setup

The content transfer experiment was conducted as a Wizard-of-Oz setup for which a web-based prototype (see Fig. 1) was developed, allowing video content to be 'transferred' between a mobile phone and a large flat screen TV connected to a computer, based on the actions of the test participant.

The user, only seeing the mobile phone and the flat screen TV, is lead to believe that (s)he actually controls on which device the video content is displayed. The user is also unaware that a wizard is observing his/her actions via live video recordings of the test scene and thereby determining on which device to show the video feed.

All participants are first introduced to a typical scenario in which they are to transfer content from their mobile phone to the TV screen. The facilitator instructs them to select, in turn, each of the four transfer methods on the home screen of the mobile phone application, and then to actually perform the transfer. The order in which the methods are evaluated is randomized for each user in order to minimize potential learning/biasing effects. Participants should comment on each of the four concepts immediately after experiencing it. Finally, after having tried the four concepts, they should indicate and justify their preferred method.

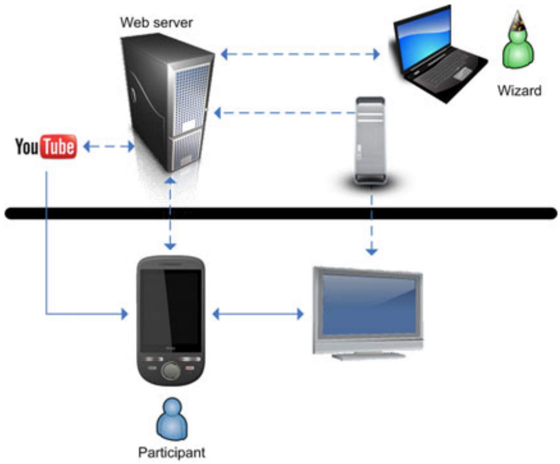


Fig. 1. Technical setup of the content transfer experiment

3.3 Results

Each participant was asked to specify a preference among the four concepts for initiating a transfer from the mobile phone to the TV. The results are shown in Fig. 2.

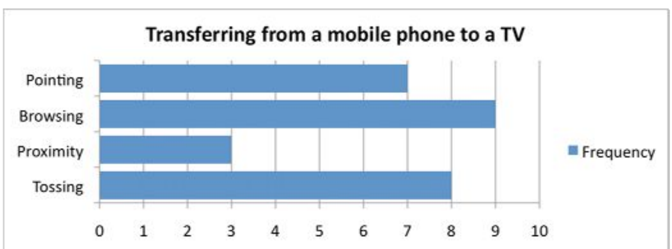


Fig. 2. Preferences when transferring content from a mobile phone to a TV

Based on these findings no significant conclusion can be drawn, although there appears to be a general dislike against the 'proximity' concept. This gives good reason to investigate closer the comments stated by the participants during the experiment. Based on these comments, the participants have been clustered into different groups indicating if they are predominantly positive or negative in their statements about each concept. From the comments collected, it can be inferred that tossing is a popular concept due to the fun of using it. Browsing and pointing are popular due to their resemblance to well-known interaction paradigms (respectively searching for devices under MS Windows and using a remote control). However proximity is not a popular concept, as it requires the user to move around physically which was perceived inconsistent with the context of watching TV.

4 Acceptability of Channel Switching Delay on a Mobile Device

To investigate the acceptability threshold of transition delay when switching between TV channels on a mobile device, video clips pre-padded with a 'transition' were used and compiled into playlists. This approach allows for full control of the delay durations without depending on network conditions and other such environmental factors. An iPod Touch was used to play the video clips and a custom-made web interface displayed on a laptop computer served to assess the transition delays.

The acceptability experiment consisted in assessing the statement "The duration of the transition was acceptable" on a 6-point Likert scale ranging from "Agree very strongly" to "Disagree very strongly". We deliberately chose a forced-choice response scale to reduce the central tendency bias. A range of transition delays (0-10 seconds) was selected based on observations of systems available today: fixed digital televisions offering short channel switching times (approximately 2 seconds) and DVB-H capable mobile phones with which longer delays are usually experienced (approximately 6-8 seconds).

In addition to identifying the threshold of perceived acceptability of the transition delays, we investigated three factors that may influence the perceived acceptability.

4.1 Possible Impacting Factors

Transition Type. Two types of transition were used between the video clips. One type consists in playing the clip of which the video is blurred while the other consists in displaying an animated icon on a blank screen. The former simulates transition conditions that allow delivering content in low quality only, while the latter simply informs the user that something is happening on the device.

Test Environment. Two environments were used as a setup for the experiment: a quiet room where nothing happened besides the test and a usability lab setup that simulated an exterior environment without actually going out in the field. In this case, the scenario for the simulation was a bus trip: the participants were sitting in a dark area, facing a video projection of a 12 minutes bus ride filmed from a 1st person view.

Video Content. Eighty-six video clips were recorded randomly from 43 Danish cable television channels during two sessions on different days. Forty playlists were then created by randomly selecting 33 different clips from the 86 available. The playlists reflect a natural browsing session throughout 33 different channels.

4.2 Results

Acceptability Threshold. Each participant experienced delay durations three times each in order to ensure data consistency. The median of the three responses is computed for each delay duration, producing an array of ratings per participant for all delay durations. Individual thresholds are then determined by the last acceptable rating when reading the array from short to long durations. This approach favors lower delay durations if an acceptable mark is given to a delay longer than the one of the first unacceptable duration. We argue that a delay rated as unacceptable should be given a higher priority, because the experiment aims at identifying the threshold at which people start getting annoyed by the delay rather than the threshold at which they stop getting annoyed by such delay. Once the personal threshold has been calculated for all participants, averaging them provides a general acceptability threshold. The first conclusion from this study is thus that the participants felt annoyed by delay durations longer than 5.7 seconds.

Effect of Factors. We then investigated the effect of the transition type, test environment and audiovisual content of the video clips on the rating of individual transition delays, according to the following hypotheses.

H1: The ratings of transition delays vary significantly depending on the type of transition used between video clips.

H2: The ratings of transition delays vary significantly depending on the test environment in which the video clips are played.

H3: The ratings of transition delays vary significantly between video clips according to their audiovisual content.

For what concerns the transition type, the one-way analysis of variance (ANOVA) performed shows that similar transition delays visualized as an animated icon were rated as more acceptable with a high level of significance ($p = 7e\text{-}4$). With regards to the test environment, the one-way ANOVA performed shows no significant level of variance between the simulated environment and the quiet room setups.

To investigate the impact of the content on the perceived acceptability of delay durations, the clips have been categorized using a collapsed version of the LSCOM-Lite content classification scheme [25] focusing on the program categories “news” or “entertainment”, the scene types “indoor” or “outdoor” and the display of a group of “people” or a single “person”. No significant effect of any of these categories was found by the ANOVA performed, which seems to indicate that the type of content does not influence the perception of transition delays.

Table 2 concludes on the three hypotheses concerning the effect of the transition type, test environment and content type on the rating of transition delays.

Table 2. Effect of three factors on the perceived acceptability of transition delays

Hypothesis	Conclusion and comments
H1	Accepted with high significance (transition delays are rated as more acceptable when illustrated with an animated waiting icon than with deteriorated content).
H2	Rejected, only a tendency: transition delays are rated as more acceptable in the lab than in the tent.
H3	Rejected, transition delays are not rated differently according to the video clip audiovisual content.

5 Conclusions and Potential Future Work

In this paper we have investigated usability issues related to ubiquitous media environments. Especially, we have focused on transferring video content from a mobile phone to a TV and on acceptable channel switching delays on a mobile device.

Among the four concepts for initiating a transfer of content from a mobile phone to a TV no significant preference has been found. The lack of conclusive result in terms of preferred method and the somewhat contradictory comments can be interpreted in two ways when it comes to inform potential designers of a market-ready application. Firstly, the pros and cons of each method may equal out among participants, who individually may value different features. In that case, the application should offer end users the possibility to choose among various transfer methods. A second solution would be to encompass all positive features found in the methods evaluated in this study: remoteness (transfer content from afar), directedness (target one specific device), memory (remember devices) and enjoyability (fun to use).

The second experiment reported in this paper shows that delays of up to 5.7 seconds are considered acceptable when switching between two TV channels on a mobile device. The type of content played did not significantly impact this result. For such delay duration, displaying a blank screen with an animated icon was perceived significantly more acceptable than displaying a blurred version of the video feed. The fact that the environment did not impact the results indicates that researchers may conduct such study in a standard usability laboratory without setting up a test environment with a high level of realism.

A real implementation of the content transfer methods would possibly uncover additional usability aspects caused by technical constraints. Additionally, examining participants' reactions when experiencing the four transfer methods could provide further insights on the preference they reported verbally. A behavioral classification scheme could be established in order to do so.

Investigating other scenarios, such as jumping directly to a known channel or skipping over several channels at once, could extend the delay study.

Acknowledgments. The work presented in this paper takes part in the CAMMP project[1], funded by the Danish National Advanced Technology Foundation. The authors would also like to thank all the participants who took part in the studies.

References

1. Thompson, M.: Putting quality first - The BBC and public space. BBC (2010), http://bbc.in/aU7tiM (February 3, 2011)
2. Berthold, A., Schmidt, J., Kirchknopf, R.: The ARD and ZDF Mediathek portals. EBU (2010)
3. Kozamernik, F.: Evolution of the BBC iPlayer. BBC (2008), http://bit.ly/bYq1Rw (February 3, 2011)
4. Oulasvirta, A., Sumari, L.: Mobile kits and laptop trays: managing multiple devices in mobile information work. In: Proc. CHI 2007, pp. 1127–1136. ACM, New York (2007)
5. Dearman, D., Pierce, J.S.: It's on my other computer!: computing with multiple devices. In: Proc. CHI 2008, pp. 767–776. AMC (2008)
6. Åkesson, M., Ihlström, C.: Towards a Ubiquitous Media Environment - Adding the e-newspaper Channel. In: Proc. TAGA (2006)
7. Buchinger, S., Kriglstein, S., Hlavacs, H.: A comprehensive view on user studies: survey and open issues for mobile TV. In: Proc. EuroITV 2009, pp. 179–188. ACM, New York (2009)

[1] http://www.cammp.aau.dk/beta/ (February 3, 2011)

8. Tsekleves, E., Whitham, R., Kondo, K., Hill, A.: Bringing the television experience to other media in the home: an ethnographic study. In: Proc. EuroITV 2009, pp. 201–210. Springer, Heidelberg (2009)
9. Harper, R., Regan, T., Izadi, S., Al Mosawi, K., Rouncefield, M., Rubens, S.: Trafficking: design for the viral exchange of TV content on mobile phones. In: Proc. MobileHCI 2007, pp. 249–256. ACM, New York (2007)
10. Park, J., Blythe, M., Monk, A., Grayson, D.: Sharable digital TV: relating ethnography to design through un-useless product suggestions. In: Ext. Abstracts CHI 2006, pp. 1199–1204. ACM, New York (2006)
11. Cesar, P., Bulterman, D., Jansen, J.: Leveraging user impact: an architecture for secondary screens usage in interactive television. Multimedia Systems 15(3), 127–142 (2009)
12. Cui, Y., Chipchase, J., Jung, Y.: Personal Television: A Qualitative Study of Mobile TV Users. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 195–204. Springer, Heidelberg (2007)
13. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1994)
14. ITU Focus Group on IPTV. FG IPTV-C-0545, Consideration on Channel Zapping Time in IPTV Performance Monitoring. ITU (2007), http://bit.ly/bVuuse (February 3, 2011)
15. Robertson, S., Wharton, C., Ashworth, C., Franzke, M.: Dual device user interface design: PDAs and interactive television. In: Proc. CHI 1996, pp. 79–86. ACM, New York (1996)
16. Hsu, C., Hefeeda, M.: Bounding switching delay in mobile TV broadcast networks. In: Proc. MMCN 2009. SPIE, Bellingham (2009)
17. Knoche, H., McCarthy, J.D.: Design requirements for mobile TV. In: Proc. MobileHCI 2005, pp. 69–76. ACM, New York (2005)
18. Kooij, R., Ahmed, K., Brunnström, K.: Perceived Quality of Channel Zapping. In: Proc. CSN 2006, pp. 155–158. Acta Press (2006)
19. ITU-T. Recommendation G.1030 - Estimating end-to-end performance in IP networks for data applications. ITU (2005), http://bit.ly/agq9ZS (February 3, 2011)
20. Godana, B.E., Kooij, R.E., Ahmed, O.K.: Impact of Advertisements during Channel Zapping on Quality of Experience. In: Proc. ICNS 2009, pp. 249–254. IEEE, Los Alamitos (2009)
21. De Wachter, J., Tytgat, D., Verstraete, V., Strobbe, M., Van Quickenborne, F., De Neve, P., De Turck, F., Demeester, P.: Optimizing channel switching for digital television through low quality streams. In: Proc. CSN 2007, pp. 55–62. Acta Press (2007)
22. Scheible, J., Ojala, T., Coulton, P.: MobiToss: a novel gesture based interface for creating and sharing mobile multimedia art on large public displays. In: Proc. MM 2008, pp. 957–960. ACM, New York (2008)
23. Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., Chin, J.: An Experimental Comparison of Physical Mobile Interaction Techniques: Touching, Pointing and Scanning. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 87–104. Springer, Heidelberg (2006)
24. Peng, C., Shen, G., Zhang, Y., Lu, S.: Point&Connect: intention-based device pairing for mobile phone users. In: Proc. MobiSys 2009, pp. 137–150. ACM, New York (2009)
25. Naphade, M.R., Kennedy, L., Kender, J.R., Chang, S.-F., Smith, J.R., Over, P., Hauptmann, A.: A Light Scale Concept Ontology for Multimedia Understanding for TRECVID 2005 (LSCOM-Lite), IBM Research Technical Report (2005), http://bit.ly/fxhlLm (February 3, 2011)

User Satisfaction of Ali Wangwang, an Instant Messenger Tool

Jie Gao and Zhenghua Zhang

UED Taobao.com, Huaxing Road 99, Xihu District,
Hangzhou, Zhejiang, P.R. China
xuezhi.gj@taobao.com

Abstract. Ali Wangwang is an instant messenger that is used by sellers and buyers to get an agreeable dealing about goods listed on Taobao.com. Buyers from Taobao.com stated that they used Ali Wangwang to communicate with sellers before almost every transaction. The objective of this study was to understand the primary factors that affect the user experience of buyers using Ali Wangwang. We designed self-reported questionnaires with questions focusing on users' overall satisfaction and evaluations of the messenger's interface and function, privacy protection, spam messages control and other related properties. We found that properties related to ease-of-use were the most significant factors in predicting user satisfaction.

Keywords: satisfaction score, instant messenger, e-commerce.

1 Introduction

According to our records, Taobao.com, founded in 2003 by the Alibaba Group, is the largest online retail platform in China with more than 370 million registered users (at the end of 2010). It provides access for businesses and individuals to open online retail stores that serve buyers across mainland China and beyond. Ali Wangwang is an instant messaging platform that was developed by the Alibaba Group. It is used by sellers and buyers to reach agreements regarding goods on Taobao.com. Based on our previous unpublished survey, approximately 99% of buyers at Taobao.com reported that they used Ali Wangwang to communicate with sellers at least once before making a purchase. More importantly, over 77% buyers reported that they used Ali Wangwang to communicate with sellers before every transaction at Taobao.com. Despite the large amount of people using Ali Wangwang, we have not systematically explored important factors that affect user satisfaction with this instant messaging platform. The objective of this study is to determine users' satisfaction with Ali Wangwang. Specifically, we would like to understand the factors that can predict the users' satisfaction scores to Ali Wangwang.

1.1 Background Information and Hypotheses

Customer satisfaction is a reflection of the degree of the positive feelings of customers towards a service provider. It is an important indicator of a customer's intention to

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 414–420, 2011.

repurchase from that provider [1]. For an online retail platform, we believe that customers' satisfaction with Taobao.com's services is influenced by the user experience of the trading website, the experience of communicating with sellers and other off-line factors such as shipping services. Here, we are only interested in exploring factors that may affect the overall satisfaction of buyers at Taobao.com when they communicate with sellers using Ali Wangwang.

Instant messenger (IM) is a convenient communication technology that is used on the Internet. A previous survey by CNNIC showed that over 70% of Chinese internet users have used at least one type of IM [2]. Most IMs provide services for friends or family members to communicate and strengthen their relationships. However, the users of Ali Wangwang are mostly buyers and sellers, and most buyers use Ali Wangwang to communicate with sellers about goods and services and to negotiate an acceptable price. Therefore, the content of communications on Ali Wangwang is mostly about goods found on Taobao.com. Previous research has suggested that consumers evaluate the online shopping environment including the online shopping website and the item display webpage [3]. We believe that the smoothness and ease of using Ali Wangwang may have a very important impact on users' overall satisfaction with their online shopping experience at Taobao.com. Our hypothesis is that Ali Wangwang's ease-of-use has a positive effect on user satisfaction scores.

The user satisfaction of general IM or mobile IM tools in China captured researchers' attention [e.g., 4, 5]; however, these results may not fully capture Ali Wangwang's special role in buyers' online shopping experience at Taobao.com. A CNNIC survey showed that when Chinese users utilized IMs, they were not only chatting with friends but also playing online games, watching online TV and performing other activities [6]. It is possible that, for IM users in China, the various entertainment functions of IMs may be as important as the chatting function. However, all previous research was based on a general IM tool; no research has discussed how users evaluate of the chatting (communication) function and non-chatting function (entertainment functions and other extended functions) of an IM in a retail context. It is possible that the chatting function, rather than the entertainment function, is of the primary importance to users' satisfaction with Ali Wangwang, because the goal of buyers with Ali Wangwang is to gain more information about the goods or services. Our hypothesis is that entertainment or other extended functions other than chatting may not influence the satisfaction of Ali Wangwang users.

Researchers have asserted that trust is a very important factor for building and maintaining relationships, and trust is also an important factor for the success of an e-commerce website [7], research showed that trust has a positive influence on a user's satisfaction with an e-commerce website [5]. Particularly, through our personal communications with users, we found that buyers sometimes did not want to share their purchasing history with other people for various reasons including embarrassment or concerns with identity theft. We believe that the protection of privacy and personal purchasing information is a crucial part of building users' trust to Taobao. Although we are aware that privacy and safety protections are very important for gaining users' trust and have put a great deal of effort into privacy and safety protections, we do not know if users have perceived our effort or whether these protections positively affect user satisfaction with Ali Wangwang. We assumed that for those users who perceived our efforts of privacy and safety protections would be more satisfied to Ali Wangwang.

2 Research Method

2.1 Questionnaire

Currently, Ali Wangwang has two versions: a buyer version, which is like a regular IM tool, and a seller version, which is more like an online call center. It enables sellers to employ multiple people for customer service through just one Ali Wangwang account. We are interested in user satisfaction with Ali Wangwang as an IM tool; therefore, our study was only concerned with buyers.

We designed a self-reported questionnaire with questions that focused on buyers' overall satisfaction and contained the following five aspects that may be relevant to daily usage of Ali Wangwang: ease-of-use, privacy and safety protections, spam messages, the number of functions used and the number of years of experience with Ali Wangwang.

To determine ease-of-use, we focused on understanding the aspects that affect the daily usage of Ali Wangwang. Ease-of-use was measured by the following six items that evaluated the most common aspects that users experienced in their daily usage: an evaluation of Ali Wangwang's interface, an evaluation of the chatting function, an evaluation for log-in speed and success rate, an evaluation of personal computer performance speed while running Ali Wangwang, and the frequency of unexpected software crashes. Each of these items was measured using a 4-point satisfaction evaluation scale ranging from extremely satisfied (4) to extremely unsatisfied (1).

As this IM tool is directly related to the online retail business, we believed that users would also be concerned with protections of privacy and safety, as well as avoiding spam messages. Spam messages were defined as irrelevant promotional information from sellers and when unknown people asked to join a user's friend list. Privacy and safety protections were also measured by the 4-point satisfaction scale. We measured the amount of spam messages by asking users for the number of spam messages they thought they had received in the past month.

Ali Wangwang provides not only a chatting service but also some convenience in functionality that is relevant to online trading (e.g., screen shots, a calculator, sending short text messages, and remote help), online gaming (e.g., casual games on Taobao.com), online financial services (e.g., paying utility fees and paying credit cards) and others. We wanted to know how many extended functions, not including communication or chatting, were used by our buyers. Therefore, we listed the extended functions and asked users to indicate functions they used other than chatting.

In addition to the services that Ali Wangwang provides, we measured the number of years of experience with Ali Wangwang. Although most users adopt Ali Wangwang for online shopping, there are some who use Ali Wangwang as a work communication tool, or for other personal purposes. Therefore, we asked users to report when they began using Ali Wangwang to communicate with sellers when making purchases in our questionnaire.

2.2 Participants

Our questionnaire was conducted using an Ali Wangwang pop-up window for seven days. Each day, we randomly selected 200,000 users and popped out a survey window while Ali Wangwang users were online. In total, 42,897 users opened our questionnaire, and 16,746 users answered the questionnaire during the survey period.

3 Results

Before any analysis was made, we checked the participants' responses to all of the questions and found that only 15,377 participants had fully completed the questionnaire. Therefore, only their responses were included in our data analysis.

The objective of this study was to understand the relationship between user satisfaction and the services provided by Ali Wangwang. Before examining this relationship, we first conducted a series of t-tests to examine whether satisfied and unsatisfied users would have the same attitude toward the items that we used the in the questionnaire. Our results showed that the satisfied users (who gave a satisfaction score of 3 or 4) gave higher scores to all aspects than the unsatisfied users (who gave a satisfaction score of 1 or 2). Because we had a very large sample size, to avoid any significant statistical differences due to the sample size, we also employed effect size to evaluate the differences between users with high and low satisfaction scores. The t-test results and effect sizes (Cohen's d) are shown in Table 1.

The largest effect size between satisfied and unsatisfied users came from the evaluation of the communication (chatting) function, followed by the number of spam messages users received and privacy and safety protections. The smallest differences were found in years of experience using Ali Wangwang and the number of extended functions used.

Table 1. The t-test results (df = 15,375) and effect sizes for the comparison of each item between satisfied and unsatisfied users

Items	*t*	Cohen's *d*
Evaluation of Ali Wangwang's interface	24.77^{***}	.15
Evaluation of Ali Wangwang's chatting function	32.51^{***}	.84
Successful log-in rate	17.26^{***}	.43
Log-in speed	12.83^{***}	.34
Personal computer performance while running Ali Wangwang	20.66^{***}	.47
Number of unexpected crashes	21.37^{***}	.48
Privacy and safety protection	18.75^{***}	.56
Number of spam messages	18.49^{***}	.63
Experience using Ali Wangwang	-3.99	.12
Number of extended functions used	-1.02	.03

*** indicates $p < .001$

The average overall satisfaction score for participants was 3.19 (out of 4) with SD = .62. We conducted a regression analysis to determine whether all aspects of the ease-of-use, privacy and safety protections, frequency of receiving spam messages, number of non-chatting functions that users tried, and years of experience of using would predict participants' satisfaction scores with Ali Wangwang with an alpha level of .05 for all analysis. The main results of the regression analysis are shown in Table 2. We found that in terms of ease-of-use, Ali Wangwang's interface, the chatting function, successful log-in rate and speed, the performance speed of personal computer while running Ali Wangwang, and the frequency of unexpected software crashes were all successful predictors for users' satisfaction with Ali Wangwang. Privacy and safety protections, the frequency of receiving spam messages, and the number of extended functions that users tried were also significant predictors of satisfaction with Ali Wangwang's services.

We did not find that the number of years of experience with Ali Wangwang was a significant predictor to overall user satisfaction scores, suggesting that although users may have used Ali Wangwang for years, they may be not all satisfied with this IM tool.

Table 2. Regression Analysis of User Satisfaction Scores for all Aspects

Items	Beta
Evaluation of Ali Wangwang's interface	.28***
Evaluation of Ali Wangwang's chatting function	.20***
Successful log-in rate	.07***
Log-in speed	.03***
Performance speed of personal computer while running Ali Wangwang	.08***
Number of unexpected crashes	.05***
Privacy and safety protections	.04***
Frequency of receiving spam messages	.04***
Years of Experience of using Ali Wangwang	-.00
Number of extended functions used	.01*

*** indicates $p < .001$, * indicates $p < .05$

4 Summary and Discussion

In our study, we examined the relationship between overall satisfaction with Ali Wangwang and users' perception of various aspects that we believed were relevant to user satisfaction.

As a communication tool that helps buyers and sellers reach agreements, the most crucial function of Ali Wangwang is chatting. Unsurprisingly, we found that satisfied users were significantly more satisfied with the chatting function than unsatisfied users. Users' evaluation of Ali Wangwang's chatting function is an important and significant predictor to users' overall satisfaction of this IM tool. We found that satisfied and unsatisfied users generally used the same number of non-chatting functions contained in Ali Wangwang. Although the number of non-chatting functions that were tried by users was also a significant predictor of user satisfaction scores, its influence on user satisfaction scores was much smaller. Our results seem to indicate that users view Ali Wangwang as a communication tool, and other functions that are not directly related to the ease of communication, such as playing online games, paying for utility fees, which do not greatly affect user satisfaction. However, we also believe that we should not totally abandon the non-chatting functions for two reasons: 1. Even though non-chatting functions were not heavily weighted by Ali Wangwang buyers, we still found that most buyers tried some of the non-chatting functions. As a product, we definitely need to keep these non-chatting functions to meet the users' calls. 2. The limitation of our current study is that we did not measure the time spent on using the non-chatting functions versus the chatting function. It is possible that users spend very limited amount of time on the non-chatting functions, and thus they cannot recall the scenarios of utilizing these functions. For future studies, we suggest to quantify the differences on usage by asking users the frequency and scenarios using both chatting and non-chatting functions.

As expected, privacy and safety protections were also significant factors affecting user satisfaction. The satisfied users stated that they trusted Ali Wangwang with privacy and safety protections and reported receiving less spam messages than the unsatisfied users. However, we speculated that the prediction power of receiving spam messages might be underestimated. We noticed that when we followed up with some users and asked about their opinions on spam message tolerance. Most of them stated that they understood that sometimes sellers need Ali Wangwang to market their online stores and goods to strangers in the same way that retailers make cold calls. Even though in a many cases, the personal interests between buyers and sellers were different. Chinese handle contradiction with compromise and tolerate it by finding a "middle way" that will not harm either side [8]. Chinese buyers do not want to receive spam messages, but their decision making process is also influenced by Chinese culture, resulting in a lower prediction power of the relationship between spam message tolerance and user satisfaction scores.

We asked users about how many years of experience they had of using Ali Wangwang, assuming that users with more experience would have higher satisfaction scores because of their loyalty to the product [5]. To our surprise, the number of years of experience using Aliwangwang was not predictive of the overall satisfaction scores. One explanation is that users with years of experience may have encountered a less polished version of this IM tool, and the past unpleasant experience may have left a strong first impression, biasing them to dislike the tool. The other possibility is that as a product, Ali Wangwang provides multiple updated versions each year. It is also possible that experienced users were still using an older version of Ali Wanwang and thus they have not experienced any of the improvements to Ali Wangwang's ease-of-use.

This study showed that, in the context of online shopping, customer satisfaction with Ali Wangwang primarily depended on factors related to ease-of-use. Our research is just exploration of how users perceive an IM tool during their online shopping. However, we do realize that user satisfaction to this IM tool is not limited to the usability factors we discussed in this short report. There are many other factors, such as, user experience with online shopping, user experience with sellers' attitude during communication, users' experience and knowledge with computers and so on, need further investigations.

References

1. Liao, C., Palvia, P., Chen, J.-L.: Information technology adoption behavior Life cycle: Toward a technology continuance theory (TCT). International Journal of Information Management 29(4), 309–320 (2009)
2. CNNIC. Twentieth statistical survey report on the internet development in China (2007), http://www.cnnic.cn
3. Demangeot, C., Broderick, A.J.: Consumer perceptions of online shopping environments: A Gestalt approach. Psychology & Marketing 27(2), 117–149 (2010)
4. Lu, Y., Zhou, T., Wang, B.: Exploring Chinese users' acceptances of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory. Computer in Human Behavior 25, 29–39 (2009)
5. Deng, Z., Lu, Y., Wei, K.K., Zhang, J.: Understanding customer satisfaction and loyalty: Am empirical study of mobile instant messages in China. International Journal of Information Management 30, 289–300 (2010)
6. CNNIC. Twentieth statistical survey report on the internet development in China (2006), http://www.cnnic.cn
7. Lee, M.K.O., Turban, E.: A trust model for consumer Internet Shopping. International Journal of Electronic Commerce 6(1), 75–91 (2001)
8. Peng, K., Nisbett, R.E.: Culture, dialectics and reasoning about contradiction. American Psychologist 54, 741–754 (1999)

Range Statistics and the Exact Modeling of Discrete Non-Gaussian Distributions on Learnability Data

Robert Hofman

Institute for Human and Machine Cognition, Pensacola, FL 32502, USA
rhoffman@ihmc.us

Abstract. A measure called i-bar is presented, which is the inverse of the mid-range derived from data on trials-to-criterion in tasks that require practice. This measure is interpreted as a conjoint measurement scale, permitting: (a) evaluation of sensitivity of the principal performance measure (which is used to set the metric for trials to criterion); (b) evaluation of the learnability of the work method (i.e. the goodness of the software tool); (c) evaluation of the resilience of the work method. It is possible to mathematically model such order statistics using negative binomial and logistic growth equations, and derive methods for generating prediction intervals. This approach involves novel ways of thinking about statistical analysis for "practical significance." The is applicable to the study of the effects of any training or intervention, including software interventions designed to improve legacy work methods and interventions that involve creating entirely new cognitive work systems.

Keywords: Range statistics, prediction intervals, rigorous usability analysis.

1 Background and Motivation

When software is introduced into the sociotechnical workplace, it represents the hypothesis that performance will be more effective and decisions will be improved (Woods, 1998). However, this hypothesis is often disconfirmed by experience. User-hostile aspects of software can induce frustration and surprise when the automation does things the worker does not understand [1]. During the development of information technology, many events result in information technology that is not human-centered—that is not usable, useful or understandable. Inadequate end-user involvement in the design process results in information technology that does not have a positive impact. Human-centering shortfalls result in a gap between the work that people must perform to achieve their primary task goals, and they work they must conduct simply because of the ways the software constrains them. User-created kluges and work-arounds are telltale signs that software is not human-centered [2].

The notorious frustrations and failures triggered by information technology interventions have led to a significant concern with evaluation and metrics in the field of software engineering and in the human factors community [3,4]. Considerable attention is being given to issues of rigor in usability analysis [e.g., 5]. Software engineers are reminded that graphical interfaces should be easy to learn, efficient to use, and

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 421–430, 2011.

subjectively pleasing. Such factors are contrasted with more traditional considerations of acquisition cost, reliability, etc.

In many information technology system development projects, usability evaluation is often based on some sort of “satisficing” criterion [6]. Users are queried about their reactions to information technology, and the sole metric is user satisfaction evaluated using surveys, questionnaires, observer ratings of user activity, or "walkthrough" interviews conducted while users attempt to conduct tasks [e.g., 7]. Sometimes the sole evaluation is whether a senior member of the organization likes the new tool, perhaps after a demonstration of design features. Researchers have recently become aware of the need to empirically distinguish between the perceived usability (or "goodness") of software and the perceived attractiveness of the interface [8]. “Learnability”—that is, the speed at which one learns to achieve primary task goals—has been proposed as a measurable variable [e.g., 9], but attempts to evaluate learnability have actually measured the readability of software documentation [10]. Both software systems engineers [e.g., 11] and cognitive systems engineers have called for new objective methods for evaluating the performance impact and learnability of software systems [4,12,13].

According to the Moving Target Law of Human-Centered Computing [14], the sociotechnical workplace is constantly changing, and this entails change in cognitive constraints. In domains of importance to society, business, and government (e.g., emergency response, military command, intelligence analysis, and many others) cognitive work mixes legacy work methods and technologies with new ones, addressing old problems while struggling with emerging challenges. Thus, procurement cannot depend solely on hierarchical task decomposition [as in 15], because the work will almost certainly have changed by the time such an analysis is complete. In fact, the more detailed the task decomposition, the more likely it will be brittle in the face of resilience tests, and the more quickly it will become obsolete. This results in what we call the *fundamental disconnect:* The time frame for effective experimentation is too slow to match the pace of change in the work and technology of sociotechnical systems.

"Effective" experimentation requires studies in which variables (e.g., display designs, software features, etc.) are manipulated and controlled. Multiple experiments are always required to peg down the causal determiners of performance and skill acquisition, especially in human-computer interaction. However, "it is difficult to sample all the things that must be sampled to make a generalization... the sheer number [of interacting factors] can lead to unwieldy research plans" [16, pp. 18-19; see also 31]. There are far too many variables to be taken into account and hence too many experimental tests to be conducted. Features of the participants (experience, intelligence, motivation, aptitude, etc.), the test scenarios (interesting, rare, easy, boring, etc.), the teams (co-located, asynchronous, dysfunctional, etc.), the tools (e.g., it is sometimes quite easy for human factors engineers to make better interfaces than ones made from a designer-centered design approach), and experimental task demand characteristics can all have causal efficacy, as can numerous mediating and moderating variables [13].

This means that an information technology development and procurement process should (by the standards of effective experimentation) hinge on a rather lengthy series of studies that go beyond measuring mere user satisfaction. Procurement would take even longer than it already does, at a time when a main challenge is to reduce procurement time [17]. By the time the relevant factors have been controlled, key variables isolated, and effect sizes estimated, design requirements will have changed and been re-evaluated, etc., the cognitive work will almost certainly have passed on to other incarnations. Therefore, we need to escape the fundamental disconnect between the time frame for experimentation and the time frame for effective change in the world of events that are a more or less unique morass of causal influences. In sum, both standard usability testing and standard controlled experimentation have limitations in regard to evaluating performance effects of technological interventions in joint human-machine cognitive work systems [13].

2 The Range Statistics Approach

In the standard view of performance evaluation, real-world variability must be restricted either by being controlled or manipulated. However, when doing the actual work, all of the variability of the world is in play. Intelligence, capability, motivation, alertness, problem difficulty, and many other factors might influence the relation between learning and performance and all are in play in real-world work environments (e.g., a command post). One worker might have high intrinsic motivation and high intelligence; another might have insufficient experience, low aptitude, and be suffering from the flu, etc. One *wants* such daunting variability of the world to be preserved during the evaluation of new technologies. We express what we call The Designer's Gamble:

> *We, the designers, believe that our new information technology is good, and that good work will result from its use. Thus, we must let the daunting variability of the world remain in the summary statistics, and we can conduct reasonable and yet risky tests of usefulness and usability. We are going to gamble that the software/tools and new work methods are so good that the cognitive work (human-system integration, etc.) will be measurably superior despite the daunting variability of the world.*

The Designer's Gamble is no fantasy on our part: proposals often promise it. Statements of the following general type often appear in proposals, which we paraphrase: "We will develop new modeling strategies for an architecture that will provide near real-time interoperability and robustness and mitigate data overload. This will then be integrated with a suite of algorithms that will automatically reconfigure the running simulation..." Such statements are promissory notes, as shown by the reliance on the word 'will.' Organizations, companies, and teams that seek to create information technology invariably base their design rationale and methodology on the Designer's Gamble.

The Designer's Gamble is an assumption made during the processes of procurement. As such, it is a leverage point for empirical analysis and, in particular, testing hypotheses about the goodness of software tools. In other words, the Designer's Gamble suggests a way around the fundamental disconnect. Range statistics represent a fast-track solution that can address questions concerning the goodness of the cognitive work, on the assumption of the Designer's Gamble. Researchers could still treat variability as something to be analyzed in exploration of hypotheses about the cognitive work, even using the familiar parametric statistical tests on data from experiments on human-computer interaction to probe the meaning of the variability (e.g., effects of co-located versus distributed teams). In other words, researchers can conduct the usual sorts of experiments they conduct to study human-computer interaction. However, at the same time, range statistics can be pulled out and utilized as a fast-track probe of the goodness of the software tools.

In the majority of experimental evaluations of human learning and task performance (including human-computer interaction), data are usually not collected during the practice trials that are part of the initial instruction. After all, the participants are just learning the basics of the task (i.e., the "button-ology"). Any data that might be collected would typically not be of any particular interest with regard to the major hypotheses being studied in the main trials (e.g., can people respond faster when provided with graphical rather than numeric representations in an interface?). We argue that performance on those earliest practice trials is a neglected resource, and a number of measures taken during practice trials could be informative. Furthermore, practice could involve training to a criterion level of performance, however many trials that might take, so that the experimenter has reason to believe that the participants have all learned the system to the same degree. This strategy serves to address the possible confound in the interpretation of the effects of the independent variables that are formative of the design of the main experiment. This same strategy can also be used to evaluate work methods through the use of range statistics. A generic design is illustrated in Figure 1.

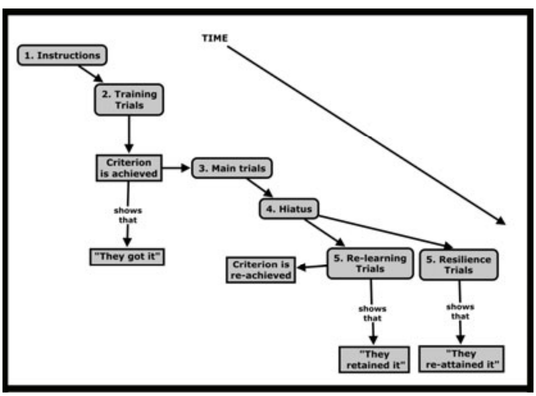


Fig. 1. A general experiment design

The derived measure of learning trials-to-criterion is familiar in experimental psychology [38]. It is used in investigations of the progress of learning, the strength of learning, and the decay of memory. It also permits control for the degree of original learning.

3 Principal Performance Measures and Range Statistics

The evaluation of performance and skill acquisition (of individuals and of teams) requires measures that reflect effectiveness and efficiency. The choice of a principal performance measure will depend on the domain, the specific job goals and other features of the cognitive work. A conceptual generic definition of a principal performance measure is *number of principal task goals successfully accomplished per unit time at work*. For example, the sensor payload operator of an unmanned aerial vehicle might have a principal performance measure of "number of targets photographed per sortie." Performance of an emergency response team in a simulated scenario might be measured by “number of victims rescued within the first hour of the response.”

A principal performance measure can be used to form the criterion for training. Participants might perform as many trials as necessary to achieve some pre-specified criterion. Practice can involve training to a lax or a strict criterion. Any reasonable value can be used initially, and whether it is liberal or conservative will depend upon the work system and the nature of the individuals who are expected to operate it. For tasks having historical performance precedent, the criterion might be based on archived data, baselines, or legacy training standards. While the performance measure and the criterion are domain- and task-specific, the derived measure of trials-to-criterion suggests a general technique for usability analysis based on examination of the "novice user's experience at the initial part of the learning curve" [9, p. 28].

The literature on the acquisition and retention of skill (both motor and cognitive tasks) encompasses hundreds of studies and review articles, and in the majority of tasks that have been studied, the criterion of “proficiency” or “minimal mastery” is typically defined as one to three errorless trials [19]. Typically, we expect a steep incline and the achievement of a reasonable level of proficiency within a short time for highly learnable systems. Counts of trials to achieve (or re-achieve) criterion are highly unlikely to have a Gaussian distribution [18]. Trials-to-criterion is an instance of a process where values are constrained by some sort of stopping rule. One would expect relatively few participants to achieve criterion on the first, second, or even perhaps third trial; if they did succeed, then one would have to conclude that the cognitive work was trivial. However, one would expect to see many participants achieve criterion after more than a few trials. Such a distribution of small numbers of small numbers typically would be highly skewed and have a “fat tail.” This characterizes distributions such as the negative binomial, illustrated in Figure 2. For such cases, an order statistic (range or median) is preferred because the average will be misleading and unrepresentative (i.e., there is a considerable difference between the mean and the median)[20]. Furthermore, we are not interested in averages; we are interested in the form of entire distributions, and especially the extremes.

We want to focus on the use of range statistics to evaluate the work method. Range statistics allow the "daunting variability of the world" to become more readily understandable through analysis.

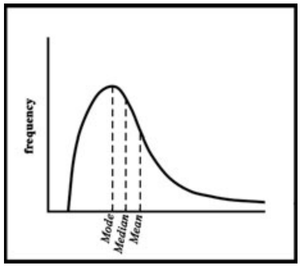


Fig. 2. A stylized distribution for trials to achieve or re-achieve criterion

4 Learnability Evaluation

Letting B stand for trials-to-criterion (or trials to re-achieve criterion) for the *best* performing participant and W stand for trials-to-criterion (or trials to re-achieve criterion) for the *worst* performing participant, the mid-range is (B + W) /2. In this case of only two numbers, the mid-range is a form of average, but we are not interested in the mid-range because it may carry with it some of the properties of the mathematical average, as interesting and useful as those properties might be. Rather, we are interested in it because of the way it conserves the performance span. The particular function of the mid-range that is of interest is the inverse of the mid-range, a new statistic that we denote with the symbol $\bar{i}$ (pronounced "i-bar"), which equals 2/(B + W). We choose this function of B and W because it is a simple transformation of the mid-range and results in numbers that fall between zero and one. The i-bar statistic is one of a class of derived measures that capitalize on the ordered nature of data. Thus, statistics in this class are also referred to as Order Statistics.

The $\bar{i}$ numerical scale can be interpreted as a conjoint measurement scale, that is, it is a measure of more than one thing. If $\bar{i}$ is quite close to 1.00, both the best and worst performing participants "got it" within just a few trials. Either the cognitive work is trivial or the criterion was set too low. If $\bar{i}$ is very close to zero, then one would conclude that the cognitive work is very difficult or the criterion was set too high. Thus, the $\bar{i}$ scale can serve as a tool for fine-tuning the criterion, or guiding the selection of the learning trials cases (or problem tasks) of an appropriate degree of difficulty.

Once one has reason to believe that the criterion is appropriate, $\bar{i}$ can be interpreted as a scale of the learnability of a work method. In this situation, if $\bar{i}$ is high, the Designer won the gamble. If $\bar{i}$ is low, the Designer lost. Details of this interpretation of $\bar{i}$ are presented in Table 1. This is just one interpretation, and is based primarily on our experience in the experimental psychology laboratory. The precise $\bar{i}$ values for the intervals are likely to be domain-specific; however, the qualitative scale and the suitability of $\bar{i}$ as a measure should apply widely.

In the Range of Stretch, the cognitive work might be extremely difficult, the work method might be very low in learnability, the criterion might have been set too high, or some combination of these may be the case. In the case of extremely difficult cognitive work, differences in $\bar{i}$ at the second decimal place might be meaningful. An example might be helicopter training, where trainees receive hours of practice at the task of hovering a helicopter, taking an average of about 20 hours to receive approval to attempt solo flight [21]. Often when people learn to fly an aircraft simulator, only after the first dozen or so practice trials, which usually result in crashes, does one begin to see trials where speed, altitude, heading, etc. are successfully maintained, even for "easy" flights.

Table 1. An interpretation of the $\bar{i}$ scale

$\bar{i}$ Scale Ranges	**Values of (B, W) $\bar{i}$**	**Desired Discrimination**
↑ Range of Trivial Cognitive Work	(1,1) 1.00 (1,2) 0.66	$\bar{i}$ between 1.00 and 0.66 suggests that the cognitive work may be trivial or that the performance criterion needs to be raised.
↓ Range of Non-trivial Cognitive Work	(1,3) 0.50	Edge of the range. Criterion may still be set too low.
↓ Range of (re)learnability	(1,4) 0.40 (2,3) 0.40 (2,4) 0.33 (2,5) 0.29 (1,6) 0.29 (3,5) 0.25 (3,6) 0.22 (4,6) 0.20	Fine discriminability is desired.
↓ Range of Stretch	(2,9) 0.18 (3,8) 0.18 (5,7) 0.16 (4,9) 0.15 ↓	Finest discriminabilty is desired.

Thus, $\bar{i}$ can be interpreted as a conjoint measure: reasonableness of the criterion, learnability, re-learnability, and stretch.

We have proposed that it is possible to augment the use of the statistic $\bar{i}$ by finding its probability distribution. This distribution depends on the choice of a probability density function to model the data (trials to achieve or re-achieve criterion). Once that is determined, the joint cumulative probability distribution function for B and W can be derived. From that, the convolution yielding the density of B + W follows and the one-to-one transformation from the density of B + W to that of $\bar{i} = 2/(B + W)$ is straightforward.

With such exact modeling, one can ask, for instance, what is the probability of an $\bar{i}$ of some value for trials to re-achieve criterion given that trials-to-criterion form a distribution of the assumed type? If that probability is low, one can conclude that the participants did not retain the original learning, and the Designer lost the gamble. If that probability is high one can conclude that the participants not only “got it” but that they also retained it, and therefore conclude that the Designer won the gamble.

5 Resilience and Team Measures

Resilience is the ability to recognize and adapt to unanticipated perturbations that might stretch the workers' competence and demand a shift of processes and strategies

[22]. For the study of resilience we can adapt a method used in clinical trials. Once there had accrued sufficient data to warrant conclusions about the learnability of a work method, there could be a session of resilience trials (rather than re-learning trials; see Figure 1) in which “the system” is stretched. This can be achieved in a variety of ways. One might simulate a communication loss, or a loss of team functionality. Performance could be evaluated by trials to re-achieve-criterion. One assumes some reasonable level of learnability but now interprets the range as a reflection of the resilience of the work system, and likewise interprets the $\bar{i}$ numerical scale as a measure of resilience.

The measure that we have described can also be applied in the analysis of work methods and technologies for teams and team cognitive work. For instance, rather than evaluating trials to achieve criterion on the part of the best (and worst) performing participants, one can evaluate $\bar{i}$ for the best (and worst) performing team. Measures can be of learnability, re-learnability and resilience with respect to the team cognitive work.

The approach we present does not seek a measure that collapses an entire set of data into a single measure of central tendency or into a single measure of variability. We are interested in testing hypotheses about extremes. "In such studies, statistical tests addressed to differences in central tendency will shield rather than reveal group differences" [23]. We remain confident that range statistics are appropriate for the analysis of the learnability of cognitive work methods. There is nothing unusual about applying extreme value statistics to practical problems, such as ranked set sampling [24].

6 Exact Mathematical Modeling of Discrete Non-Gaussian Distributions

A theoretical foundation for Range Statistics will involve exact mathematical modeling of discrete non-Gaussian distributions and some method for deriving probabilities and testing hypotheses. Both theory and empirical experience imply that the distributions of i-bar will be highly non-Gaussian. Therefore, a premise in the investigation of range statistics is that it is not appropriate to rely on the mean and variance and the analysis of variability by averaging procedures. The analysis of range statistics looks more closely at the median and mode in relation to the range. In such contexts as training, the range is especially interesting because best performance is a benchmark and worst performance represents a potential training problem. Thus, the range itself is of inherent interest.

Close-to-exact fits would be possible and would be preferred to approximations for purposes of statistical analysis. Although the standard approach in mathematical psychology is to build formulae based on a theory of learning, much can be learned from using an empirically determined probability model for the data. Once this is set, a probabilistic structure can be created. Very little research exists on the possible probability density functions of trials-to-criteria data. That question is by itself of interest, and experiments generating such data would be quite useful. What we do know is that the probability density function should be steep on the left and have a fat tail on the right. The natural occurrence of distributions that take the form of the negative binomial, and the applicability of modeling the negative binomial, have been pointed out in the psychometrics literature [25]. Other probability density functions incorporating the degree of learning achieved from trial to trial show promise as well.

There is reason to believe that trials-to-criterion will take the form of the negative binomial [26]. A second family of distributions that might fit to trials-to-criterion data is based on the logistic growth model. This is a cumulative distribution function suited to describing processes in which there is growth up to some limit, that is, the S-curve growth of some set. A third model, specified conditional probabilities, would assign conditional probabilities describing the probability of success, *given* that successive failures have occurred.

Regardless of the method of determining the probability density function for the data, the probability density function for i-bar can be calculated. From that it is possible to derive what in classical statistics are called *confidence or prediction intervals,* the probability that i-bar will fall within certain ranges of values. From these intervals statistical inferences can be made. For example, one can ask, "Given the observed data on original learning (trials-to-criterion) of B = 3 and mid-range = 2, what is the probability that a team would re-achieve criterion on 14 trials as the worst performance in the resilience test?" From that probability, and the $\bar{i}$ result, one might derive conclusions concerning the importance of adding into the work method some means for coping with the particular resilience factors that were examined in the resilience trials.

Our progress to date shows that questions about the likelihoods of values of range statistics on discrete non-Gaussian distributions can be framed mathematically and answered. A worked-out statistical theory for the derived measure of trials-to-criterion does not exist. Such a theory, or an appropriate numerical substitute, is within our grasp; as is a software tool to support modeling, range statistics analysis, and the generation of prediction intervals.

References

1. Hoffman, R.R., Marx, M., Hancock, P.A.: Metrics, metrics, metrics: Negative hedonicity. IEEE: Intelligent Systems, 69–73 (March-April 2008)
2. Koopman, P., & Hoffman, R.R.: Work-Arounds, Make-Work, and Kludges. IEEE: Intelligent Systems, 70–75 (November/December 2003)
3. Goguen, J.: Towards a social, ethical theory of information. In: Bowker, G., Gasser, L., Star, L., Turner, W. (eds.) Social Science Research, Technical Systems, and Cooperative Work, pp. 27–56. Lawrence Erlbaum Associates, Hillsdale (1997)
4. Neville, K., Hoffman, R.R., Linde, C., Elm, W.C., Fowlkes, J.: The procurement woes revisited. IEEE Intelligent Systems, 72–75 (January/February 2008)
5. Rosson, M.B., Carroll, J.M.: Usability engineering: Scenario-based development of human computer interaction. Morgan Kaufman, San Francisco (2002)
6. Gillan, D.J., Bias, R.G.: Usability science I: Foundations. International Journal of Human-Computer Interaction 13, 351–372 (2002)
7. Sears, A.: Heuristic walkthroughs: finding the problems without noise. International Journal of Human-Computer Interaction 9, 213–234 (1997)
8. Hassenzahl, M., Monk, A.: The inference of perceived usability from beauty. Human-Computer Interaction 25, 235–260 (2010)
9. Nielsen, J.: Usability engineering. Academic Press, San Diego (1993)

10. Kantola, N., Jokela, T.: Determining high-level quantitative usability requirements: A case study. Springer, Berlin (2007)
11. Rubin, J., Chisnell, D.: Handbook of usability testing, 2nd edn. Wiley, Indianapolis (2008)
12. Hoffman, R.R., Neville, K.N., Fowlkes, J.: Using cognitive task analysis to explore issues in the procurement of intelligent decision support systems. Cognition, Technology, and Work 11, 57–70 (2009)
13. Roth, E.N., Eggleston, R.G.: Forging new evaluation paradigms: Beyond statistical generalization. In: Patterson, E., Miler, J. (eds.) Macrocognition Metrics and Scenarios, pp. 204–219. Ashgate, London (2010)
14. Ballas, J.A.: Human centered computing for tactical weather forecasting: An example of the Moving Target Rule. In: Hoffman, R.R. (ed.) Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making, pp. 317–326. Erlbaum, Mahwah (2007)
15. Shepherd, A.: Hierarchical task analysis. Taylor and Francis, London (2001)
16. Firestone, W.A.: Alternative arguments for generalizing from data as applied to qualitative research. Educational Researcher 22, 16–23 (1993)
17. Public Law 111-23-Weapon Systems Acquisition Reform Act of (2009)
18. Woodworth, R.S.: Experimental psychology. Henry Holt and Company, New York (1938)
19. Hoffman, R.R.: Accelerated Proficiency and Facilitated Retention: Recommendations Based on An Integration of Research and Findings from a Working Meeting. Report on Grant FA8650-09-2-6033, U.S. Air Force Research Laboratory, Wright-Patterson AFB OH (2010)
20. Newell, K.M., Hancock, P.A.: Forgotten moments: Skewness and kurtosis are influential factors in inferences extrapolated from response distributions. Journal of Motor Behavior 16, 320–335 (1984)
21. Still, D.L., Temme, L.A.: Configuring desktop helicopter simulation for research. Aviation Space and Environmental Medicine 77, 323 (2006)
22. Woods, D.D.: Engineering organizational resilience to enhance safety. Presentation at the Eighth International Conference on Naturalistic Decision Making, Asilomar Conference Center, Monterey, CA (June 2007)
23. Siegel, S.: Non-parametric statistics for the behavioral sciences. McGraw Hill Book Company, New York (1956)
24. Amin, R.W., Li, K.: The EWMA maxmin tolerance limits. The International Journal of Quality and Reliability Management 17, 27–41 (2000)
25. Patil, G.P.: On the evaluation of the negative binomial distribution with examples. Technometrics 2, 501–505 (1960)
26. Hill, L.B., Rejall, A.E., Thorndike, E.L.: Practice in the case of typewriting. Pedagogical Seminary 20, 516–529 (1913)

Measuring Cultural Markers in Arabic Government Websites Using Hofstede's Cultural Dimensions

Nouf Khashman and Andrew Large

Shool of Information Studies, McGill University,
3661 Peel, Montreal, QC. H3A 1X1, Canada
nouf.khashman@mail.mcgill.ca, andrew.large@mcgill.ca

Abstract. This study examines the design characteristics of government web interfaces from three Arab countries using Hofstede's cultural dimensions. Organizational and graphical elements from 30 ministry websites from Egypt, Lebanon and Saudi Arabia were examined using content analysis. Element frequency scores were correlated with Hofstede's dimensions and interpreted based mainly on the model developed by Marcus and Gould. The results suggest that Hofstede's model of culture does not fully reflect the design characteristics of Arabic interfaces.

Keywords: Arab countries, Culture, Hofstede, Web design, Government websites.

1 Introduction

Culture is argued to be one of the attributes affecting the usefulness and usability of interfaces [14]. As such, researchers have investigated how different cultures represent themselves on the web, and how culture affects website design and usability. The rationale is that by localizing an interface through the incorporation of culturally appropriate design features, an interface becomes both more attractive and more functional for its users [11]. The bulk of the research in this domain has employed Hofstede's model [5, 6], based on the interpretations made by, amongst other researchers, Marcus and Gould [9].

2 Hofstede's Cultural Dimensions

Since their first introduction in 1980, Hofstede's cultural dimensions have had a profound influence on the development of cross-cultural studies in the social sciences [15]. Based on data collected from IBM employees from 40 countries, Hofstede proposed four dimensions of culture which differentiate between national cultures based on comparative scores assigned to individual countries. These dimensions comprise: *Power Distance*-the extent to which the less powerful members of organizations expect and accept that power is distributed unequally; *Individualism*-the extent to which individuals are integrated into groups; *Masculinity*-assertiveness and competitiveness

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 431–439, 2011.

versus modesty and caring; and *Uncertainty Avoidance*-intolerance for uncertainty and ambiguity. A fifth dimension, *Long-Term Orientation*-the degree of future orientation-was added later in 1982 when Hofstede expanded his model to include 10 more individual countries and three regions. For these regions, Hofstede grouped together several countries based on the assumption that they have similar cultural traits. These regions are East Africa, West Africa, and the Arabic-speaking countries [6]. In the latter he included seven countries: Libya, Iraq, Kuwait, the United Arab Emirates, Egypt, Lebanon and Saudi Arabia.

The Arabic-speaking countries had been initially surveyed both in 1969 and 1972, but when Hofstede tried to extend the country list in 1982, he found that IBM had accidentally wiped the tape containing the raw survey data. The only data that were saved pertained to the total region, so Hofstede was forced to treat these countries as one region, whereas he might have wanted to keep at least Egypt and Lebanon separate as Hofstede himself [6, p. 52] admitted that this made the region culturally less homogeneous than would be desirable.

As a group, these seven countries scored high on the Power Distance (80), Uncertainty Avoidance (68), and Masculinity (52) dimensions, while scoring low on the Individualism (38) dimension. The only dimension that does not have any scores for these countries is Long-/Short-Term Orientation.

In web design, interfaces with high Power Distance have an emphasis on social models such as nationalism and religion, have a strong focus on authority where prominence is given to leaders [9], and have restricted information access [1], and vice versa for low Power Distance. Interfaces with high Individualism (i.e. low Collectivism) are customizable, have images of individuals rather than groups [1], and motivation is based on personal achievements [9]. Masculinity in interfaces can be reflected through traditional gender distinctions between users, navigation is oriented to exploration and control, graphics and animation are used for utilitarian purposes [9], and they tend to have a visitor counter [13]. Feminine interfaces on the other hand have blurred gender roles, where tasks are accomplished through mutual cooperation and attention is gained through visual aesthetics. Interfaces with high Uncertainty Avoidance have a simple design with limited choices and a restricted amount of data, and vice versa for low Uncertainty Avoidance interfaces. Finally, users of interfaces with Long-Term Orientation must have patience in order to achieve results and goals, while users of interfaces with Short-Term Orientation have the desire for immediate results and achievements of goals.

3 Arab Countries in Cultural Design Studies

Cultural similarities and differences in web design have been discussed at length in the literature. However, Arab countries have received limited attention in this research area, despite the fact that 19% of the overall Arab population uses the internet [7] and therefore could potentially benefit from this research. Some studies that did opt to include web pages from Arab countries limited their country choices to as few as two countries. For example, in a study that included systematic inspection of design elements that are possibly preferred within a particular cultural group, Barber and

Badre [2] selected websites only from Lebanon and Saudi Arabia, which were initially chosen by Hofstede himself. Their findings indicate that these websites had a high frequency of right-to-left orientation and high frequency of flags in the government genre, relating to Uncertainty Avoidance and Long-Term Orientation dimensions respectively.

Zahir, Dobbing, and Hunter [17] selected national web portals also from two countries, Egypt and Morocco. Their results showed that websites from Egypt had a strong focus on the Egyptian culture, reflecting a high Power Distance characteristic. While websites from Morocco had a good presentation of women's issues and non-Islamic reference, relating to the Masculinity and Power Distance dimensions respectively.

Callahan [3], on the other hand, analyzed a total of 20 interfaces from the group of seven Arab countries included in Hofstede's model in her study of cross-cultural differences in the design of university websites. Although the number of websites from each country was not specified, the results pertaining to the Arab countries overall suggest that most of the design elements on their interfaces match their description on Hofstede's cultural dimensions.

In a study conducted by Marcus and Hamoodi [10], the researchers analyzed Arabic educational websites from Jordan, Egypt, and the United Arab Emirates aiming to determine whether or not the websites reflect Arabic culture. The results of this study show again that most of the design elements on these interfaces correspond to their characteristics on Hofstede's cultural dimensions.

4 Methodology

4.1 Country Selection

It was first necessary to define an "Arabic country", as different definitions will produce different member states falling within this category. As part of an ongoing research that includes all Arabic-speaking countries, a random three from Hofstede's original group of seven were included in this study: Egypt, Lebanon and Saudi Arabia (their home pages are reproduced in Figures 1-3).

4.2 Government Genre

Similar to Barber and Badre [2] and Cyr and Trevor-Smith [4], websites from the government genre were chosen for analysis. Government websites provide a sufficiently large sample size across the countries. In addition, these websites are created as a means of interaction with the locals; therefore they are presumably intended for a particular culture or nation, rather than the worldwide Internet community. Moreover, it is expected that designers who belong to the local culture created these websites, consequently likely reflecting the socio-cultural, technological and economic characteristics of their intended cultures in order to be successful in the services they provide.

The sample frame was based on the lists of government websites provided on the web portals of the countries' governing body. A simple random sampling was conducted to determine the final sample to be included in this study.

4.3 Analysis

The home page is argued to be the most important page of any website [12]; given its attention-grabbing and organizational roles, it is likely to contain many central elements of Web design [16]. Therefore, we applied content analysis method to the Arabic home page of 30 websites of ministries (except when the website was only provided in another language) from the selected three countries, 10 from each.

According to Krippendorff [8], content analysis is a valid method used to describe trends in a communication context, allowing researchers to draw inferences on patterns and differences among similar components of that communication context. The components in this research are the web design elements, described by Barber and Badre [2] as “cultural markers”, which have been argued to be prevalent and possibly preferred within a particular cultural group. The analysis focused on the graphical, organizational, and navigational elements that consist of: presence of social models (national, religion, etc.), restriction of information access, logo depiction (traditional vs. modern), presence of animated images, page directionality (vertical vs. horizontal), number of hyperlinks, menus (simple vs. complex), presence of customization (font size, color, etc.), presence of a visitor counter, search engine, site map, and FAQ links. Images of people were analyzed according to the number (single vs. group), gender (male vs. female vs. mix), status (leader vs. citizen vs. mix). The difference of color choices and language options were also recorded.

Statistical analyses were performed using SPSS program based on the specific level of measurement for each variable. Categorical variables (e.g. presence of a search engine) were counted, converted to a percentage, and compared using non-parametric chi-square test. Descriptive statistics were used to describe continuous variables (e.g. number of links). The results then were compared to the description of Arab countries on each of Hofstede’s cultural dimensions based on the interpretation in the literature [1, 9, 13].

4.3 Reliability

Two coders, fluent in Arabic and English, were trained in the coding scheme on non sampled websites. The proficiency of the Arabic language proved useful to determine the status of persons in images (official or citizen), especially when these images were associated with news items. The coders then analyzed the home page of each website to count the occurrence or non-occurrence of each of the cultural markers. An overall acceptable inter-coder reliability of 81% was established using Krippendorff's alpha.

5 Results

The content analysis of Arabic government web interfaces from Egypt, Lebanon, and Saudi Arabia suggest that overall there were significant differences in the depiction of cultural markers between the websites of these three countries on the one hand and the characteristics of Arabic-speaking countries on Hofstede’s dimensions on the other. The chi square results for categorical variables are shown in Table 1, followed by descriptive statistics of the continuous variables, in addition to variables that are not necessarily associated with a particular dimension.

Table 1. Results of Chi-square test for the design elements

Dimension	Associated design element	%	χ^2	*df*	***p***
Power Distance	Social models	30%	4.8	1	.028
	Restriction to access	13%	16.13	1	.000
	Images of leaders	39%	ns	2	
Collectivism	Customization	23%	8.53	1	.003
	Images of groups	40%	4.70	1	.029
	Images of leaders	39%	ns	2	
Masculinity	Visitor counter	20%	10.80	1	.001
	Images of men	68%	49.77	2	.000
	Animated images	70%	4.80	1	.028
Uncertainty Avoidance	Vertical directionality	80%	10.80	1	.001
	Simple menus	37%	ns	1	
Long-Term Orientation	Search engine	70%	4.80	1	.028
	Site map	53%	ns	1	
	FAQ	23%	8.53	1	.003

Fig. 1. Ministry of Health, Egypt

Fig. 2. Ministry of Health, Lebanon

Fig. 3. Ministry of Health, Saudi Arabia

Additionally, Arabic is the main language in 83% of the websites (20% of which have no secondary language), and is a secondary language in 13% of them.

English is the second most used language with 17% of the websites using it as a main language and 63% using it as a secondary one. French is also used as a secondary language in 27% of the websites, mostly along with English.

As for background colors, 63% of the websites use white, 20% use blue, 17% use other colors (e.g. beige, brown, green). Interestingly, dominant colors are mostly blue with 40%, white with 23%, green with 20%, and 17% for other colors (yellow, grey, brown, ect.). Also the number of links varied greatly from one site to another, ranging from 15 to 240.00 links, with a total number of 2561.00 (M= 85.37, SD= 48.57).

6 Discussion and Conclusion

The preliminary results suggest that Hofstede's model of culture does not fully reflect the design characteristics of Arabic interfaces.

As countries with a high score on Hofstede's Power Distance dimension, we would expect Arab countries to have frequent use of social models, restrictions to access information, images of leaders, and traditional logos. Presence of social models and restriction of information access were less than expected, while the number of images of leaders was not significantly different than those that had mix status or citizens alone. Logo depiction did not have an acceptable agreement between coders, and therefore it was not reported.

As belonging to a collectivist culture in Hofstede's model, Arabic interfaces would be expected to show low customization (i.e. font size, color, etc.), have more frequent images of groups and images of leaders than individuals and "ordinary" citizens. This assumption was confirmed for customization, but not for images of leaders or images of groups.

Arab countries are described as having a relatively masculine culture in Hofstede's model. Therefore, we would expect their interfaces to have frequent use of visitor counters, images of men and animated images. This assumption was confirmed for images of men and the presence of animated images, but not for visitor counter.

Scoring low on Hofstede's Uncertainty Avoidance, we would expect Arab countries to have more frequent horizontal pages, simple menus, and relatively low number of links. This assumption was confirmed for the last two elements.

Although Arab countries do not have a score on Hofstede's fifth, Long- vs. Short-Term dimension, they could be described as long-term oriented. Therefore, it was interesting to find that most websites have search engines and site maps as means to search for information, contrary to what we would expect. But as expected, these websites have few FAQs links.

While the results of this study confirm the results of other studies that included Arab countries [2, 17], in regard to the use of culturally favored colors and depiction of images, they also refute the results of other studies, for example, Callahan [3], in regards to the presence of search engines and page orientation. The differences in the results could be attributed to the difference in country selection or website genre, therefore further research is needed.

This study has several limitations. First, websites from only three countries out of a possible seven that Hofstede included in his model were analyzed, limiting the comparison between countries. Second, the results might be influenced by the type of website chosen for analysis, as noted by Barber and Badre [2]. Third, we conducted a quantitative analysis on the design elements. The research would be more robust and representative when qualitative analysis is included to investigate how people in these countries interpret and respond to the design of their government web interfaces.

The study results and limitations imply the need for further research. Future investigation will focus on the remainder of the countries from the group of seven, as well as other Arab countries excluded from Hofstede's model. In addition, websites from another genre than government will be included in order to yield more meaningful results and to produce valid comparisons across genres. The much wider question remains, of course, as to whether the usability of an Arabic website is enhanced by designing it in accordance with these cultural markers.

References

1. Ackerman, S.: Mapping User Interface Design to Culture Dimensions. Paper presented at the International Workshop on Internationalization of Products and Systems, Austin TX (2002), http://www.usj.edu.lb/moodle/stephane.bazan/obs_interculturelle/culture%20dimensions%20in%20WS.pdf (retrieved April 7, 2010)
2. Barber, W., Badre, A.: Culturability: the merging of culture and usability. In: 4th Conference on Huamn Factors and the Web (1998), http://research.microsoft.com/users/marycz/hfweb98/barber/ (retrieved September 24, 2008)
3. Callahan, E.: Cultural Differences in the Design of Human-Computer Interfaces: A Multinational Study of University Websites. Published thesis, Indiana University (2007)
4. Cyr, D., Trevor-Smith, H.: Localization of Web Design: An Empirical Comparison of German, Japanese, and United States Web Site Characteristics. Journal of the American Society for Information Science and Technology 55(13), 1199–1208 (2004)
5. Hofstede, G.: Culture's Consequences: International Differences in Work-Related Values. Sage Publications, Beverly Hills (1980)
6. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications, Thousand Oaks (2001)
7. Internet World Stats, http://www.internetworldstats.com/ (accessed on August 1, 2010)
8. Krippendorff, K.: Content Analysis: An Introduction to its Methodology. Sage Publications, Beverly Hills (2004)
9. Marcus, A., Gould, E.: Cultural Dimensions and Global Web User-Interface Design: What? So What? Now What? In: Proceedings of the 6th Conference on Human Factors and the Web. Austin, Texas (June 2000), http://www.amanda.com/resources/hfweb2000/hfweb00.marcus.html (retrieved September 25, 2008)
10. Marcus, A., Hamoodi, S.: The Impact of Culture on the Design of Arabic Websites. In: Aykin, N. (ed.) IDGD 2009. LNCS, vol. 5623, pp. 386–394. Springer, Heidelberg (2009)
11. Nielsen, J.: International Use: Serving a Global Audience. In: Designing Web Usability, pp. 315–344. New Riders, Indianapolis (2000)

12. Nielsen, J., Tahir, M.: Home Page Usability: 50 Websites Deconstructed. New Riders, Indianapolis (2002)
13. Robbins, S., Stylianou, A.: A Study of Cultural Differences in Global Corporate Web Sites. The Journal of Computer Information Systems 42(2), 3–9 (2002)
14. Shneiderman, B., Plaisant, C.: Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley, Reading (2010)
15. Søndergaard, M.: Research Note: Hofstede's Consequences: A Study of Reviews, Citations and Replications. Organization Studies 15(3), 447–456 (1994)
16. Weare, C., Lin, W.: Content Analysis of the World Wide Web: Opportunities and Challenges. Social Science Computer Review 18, 272–292 (2000)
17. Zahir, S., Dobing, B., Hunter, G.: Cross-Cultural Dimensions of Internet Portals. Internet Research 12(3), 210–220 (2002)

Different UI, Same UX: A Design Concept for Implementing a Locally-Optimized and Globally-Unified User Experience

Sung Woo Kim[1], Han Kyung Jo[2], and Da Yun Ha[3]

[1] KT R&D Center, 17 Woomyun-dong, Seocho-gu, Seoul, South Korea
[2] ITS Business, Samsung Electronics, Seocho-gu, Seoul, South Korea
[3] Interaction Design, Graduate School of Techno Design, Kookmin University, 861-1, Jeongneung-dong, Seongbuk-gu, Seoul, South Korea
caerang@gmail.com, hankyung.jo@gmail.com, yahadesign@gmail.com

Abstract. "Different UI, Same UX" is a design concept originated from UX research on content and information services in multi-screens, also known as 3-Screens. The biggest UX challenge in such an environment is that it is composed of different devices in different interfaces that need to work together to provide one integrated service. In a heterogeneous environment like this, "Different UI" emphasizes creating UIs that fit each device and therefore put less weight on consistency. At the same time, "Same UX" highlights to need to maintain one coherent branded UX identity across devices; coherence being a higher-level of consistency. This paper introduces how these two priorities can be reconciled into one design concept. The paper elaborates on the definition of "Different UI, Same UX" with a number of baseline ideas. Several industry examples that we believe illustrate this concept are also discussed.

Keywords: branding, coherence, consistency, design concept, 3-Screens, user experience.

1 Introduction: What Is "Different UI, Same UX"?

"Different UI, Same UX" is a design concept useful to heterogeneous environments such as 3-Screens, where different kinds of devices provide unified information and content services [14]. It aims to achieve good usability on each device and, at the same time, one overall user experience resonating with corporate branding. In terms of design practice, it's a concept to encourage UX designers to design different user interfaces that match to each device and, at the same time, to put effort into creating one branded user experience [9].

"Different UI" refers to user interface that is optimized to its device. For example, user interface consistency may be tightly controlled within a device, but loosely controlled across devices. In other words, UI elements should be optimized and consistent within the scope of individual device.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 440–448, 2011.

"Same UX" focuses on tailoring experience elements among devices in order to offer one user experience that delivers on the organization's brand promise. Here, the emphasis is not on traditional concept of alignment but rather on making a coherent experience story [15]. Hence, from the "Same UX" perspective, coherence counts more than consistency. This is critical in 3-Screens services where a number of devices cooperate together to provide one service [16][17]. "Same UX" is a higher-level design technique to create experiential homogeneity out of totally different user interfaces.

2 Research Background

2.1 "Different UI, Same UX" in the Era of 3-Screens

The research wad derived from handling user experience challenges in multi-screen information and content service in telecommunication, broadly known as 3-Screens (figure 1.)

Fig. 1. 3-Screens: originally conceptualized at AT&T where content is shared among PC, TV and mobile phone. Nowadays the number "three" has become figurative to mean multiple screens.

Companies like AT&T or KT (Korea Telecom) generally do not manufacture their own devices but buy them from IT manufacturers like Samsung or Apple to set up their multi-screen services. Consequently their services easily become heterogeneous where different devices with different user interfaces are mixed to provide common services [5][6][11]. In such cases, user experience designers are often faced with the problem of optimizing usability quality of each different device while providing same user interface for brand identity purposes [1][9][12][13]. It becomes more challenging when UI alignment and standardization are treated as a critical user experience issue.

2.2 Application to Other Design Domains

While it is true that 3-Screens is the main environment for this research, "Different UI, Same UX" can be applied to more general UX standardization efforts. Manufacturers such as Samsung, LG, and Apple have multiple product categories, including laptop, monitor, printer, mobile phone, TV, etc. Demand from marketers for alignment and standardization in both physical and screen UI is again a typical challenge given to UX designers in these companies.

We have seen many design failures in these situations where design has been carried out with a rigid and inflexible approach to consistency; that is, taking consistency as the most critical criterion to achieve alignment and standardization [10]. Pushing for consistency in a heterogeneous device environment has the potential risk of poor usability because every design decision must suit all the way down to the lowest device in the environment; an issue generally referred to as "matching to lowest common denominator."

3 Baseline Ideas

Below are the ideas that form the baseline of "Different UI, Same UX."

3.1 UX ≠ UI

The very fundamental idea of "Different UI, Same UX" lies in the definition of UX as "beyond UI and not equivalent to it." Obviously, if we assume "UI = UX" then the idea of "make UI different but make UX same" can't exist. The UI in "Different UI, Same UX" refers to traditional user interface such as font size, color, navigation style, UI layout, IA, interaction logic and flow, etc.

In this sense, user experience is a set of experience elements at a bigger scale [3]. UI is considered to be a portion of it; in fact, UI can be named as an interaction experience element. There are other experiential elements that comprise user experience, depending on what we design. It is UX designer's role to identify these elements [4].

3.2 Levels of UX Standardization

Figure 2 shows the 4 levels of UX standardization. This idea, along with "Different UI, Same UX," is another outcome of our UX research for 3-Screens.

Simply put, elements fall under Level 1 and 2 are traditional user interface elements and therefore subjects of "Different UI."[1] The first level is all about styling (color, font, etc.) The second level concerns alignment in interaction rules and logics. Typical examples are date format (e.g.: dd/mm/yy) and navigation style (e.g.: all list browsing must use jumping highlight.) Many organizations put effort to standardize their UI around these levels.

Level 3 is where we begin to consider user experience elements other than traditional UI elements. All experiential elements and higher-level interaction- experience elements that are hard to park into level 1 and 2 are placed under level 3. Level 4 is a stage to build UX ecosystem. Level 3 and 4 are higher level UX standardization stages recently arisen as IT paradigm shifts from standalone devices and services to network, ubiquitous, and cloud based devices and services. They are the ones to apply "Same UX"; particularly level 3.

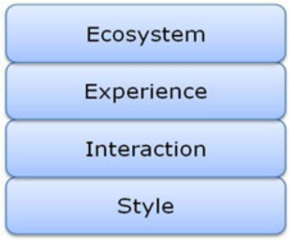


Fig. 2. The 4 Levels of UX Standardization

3.3 Consistency vs. Coherence

Consistency is an undisputable core principle in user interface design [7]. However, we do need to be cautious in applying consistency in new heterogeneous environment. While sticking to consistency is still valuable in UI per device, rigid deployment of consistency among devices may result in poor quality. [1]

For example, a menu navigation style optimized to a remote control with four-directional keys may not fit to a device with touch-screen interaction. Here, an obsessive pursuit of consistency will force UX designers to create a navigation style that works on all devices but are not optimized to any of them, resulting in the "lowest common denominator" mentioned in 2.2. A wise approach is to allow designers to create customized navigation style for each device; the practice of "Different UI."

However, this does not imply that consistency is of no concern in "Different UI, Same UX." UX designers should look into higher level of consistency in order to realize "Same UX." Coherence is the new name for this high-level consistency.[8][17] Coherence, according to Collins-COBUILD English Dictionary, means "a state of situation in which all the parts or ideas fit together well so that they form a united whole." Focusing on the two underlined phrases of this definition, our interpretations in terms of user experience design are:

- *fit together well*: elements are not identical to one another and they keep their own characteristics but they go well smoothly that anyone can tell they are tightly connected.
- *form a united whole*: thanks to their tight connection, elements are seamlessly integrated to bring one experience.

Metaphorically speaking, a UX with good coherence is like a good novel where introduction, development, turn and conclusion are seamlessly united to deliver a well organized story [15]. Obviously we do not see same text repeated in each phase of a story, which would be redundant and non-sensical.

3.4 Locally Optimized Consistent UI & Globally Unified Coherent UX

Based on the description from 3.3, "Different UI, Same UX" can be rephrased to "locally optimized consistent UI & globally unified coherent UX." As mentioned above, consistency still counts for providing optimized usability per device. For cross-device user experience, coherence is the key to realizing a globally unified user experience among devices [2][17]. This is the core concept of "Different UI, Same UX."

4 Case Studies

In this chapter we introduce several cases from industry as design examples of "Different UI, Same UX."

4.1 Netflix -Same Application Multi Device

Netflix is a popular DVD-rental service in the USA, which has expanded its business into VOD service. Many home entertainment devices like game consoles and connected TVs carry Netflix as their built-in VOD services.

Neflix UIs in each device are designed in accordance with their devices' local user interfaces and input methods. They offer familiarity to device users therefore enable seamless experience in overall sense. Such design approach has resulted in various different versions of Netflix UI with different look-and-feels per device.

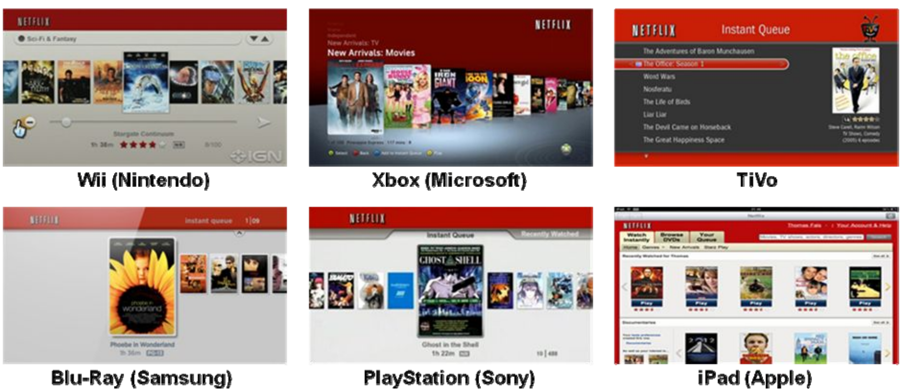


Fig. 3. Netflix service offered in various devices including Nintendo Wii, Microsoft Xbox, TiVO, Samsung blu-ray player, Sony PlayStation3, Apple iPad, etc. (http://www.netflix.com).

As shown in figure 3, Netflix VOD in Nintendo Wii displays a simplistic rotational list optimized to its gesture-based remote control. It provides arrows and sliders for users to easily interact with the list. Netflix in Microsoft Xbox also uses coverflow UI similar to Xbox's local content list.

The design approach we see from Netflix is a good example of "Different UI" such that each Netflix UI maintains consistency with its device's own UI. Consequently, device manufacturers can provide comfortable and familiar UI to their customers with relatively less effort in integration. Netflix also minimizes its own consistency to visual identities like logo and brand color, which enables well-designed locally optimized consistent UI. On the other hand Netflix provides same genre categorization and content among devices, successfully realizing same UX.

4.2 PUI (Physical User Interface)

In the case of manufacturers with wide ranges of product, iconic PUI elements are made specific to each product. PUI elements can be applied within a product group or across various product groups. An interesting thing about iconic PUI elements is that when it's applied to a product group it enhances usability through consistency, and when it's applied to different groups it brings coherence to different products that enhance brand identity.

As such, the use of iconic PUI elements is a design application that lies somewhat between "Same UI, Same UX" and "Different UI, Same UX," depending on the degree of resemblance of iconic PUI elements applied across different product groups.

The typical Iconic PUI includes the following: 1) Blue LED such as HP Printer, PC and Serve applied in various devices, 2) Touch Screen angle in HP PhotoSmart Printer Series, 3) Power Button of Sony laptop computer, 4) Sony Keyboard skin.

Fig. 4. HP–Printer, PC, Server(www.hp.com), Sony- Keyboard Skin, Laptops(www.sony.com)

4.3 Apple

Apple brands its user experience by maintaining a certain level of continuity as their products evolve, which naturally leads to "same UX" over "different UI" throughout product generations. A good example is Apple iPod's scroll/flick interaction constantly provided over different input mechanisms from its first to later generations.

The first generation of iPod had "physical scroll wheel" that spun mechanically. The second generation iPod was equipped with "touch-sensor wheel" and nowadays iPod Touch in "flicking on touch screen" without any wheel. While hardware control mechanism changed dramatically Apple kept its controlling experience to what we can refer now as "flick."

The similarity between scroll and flick is that both trade off accuracy for emotional and entertaining experience by providing motion effects such as acceleration and gradual-stop. Flick inherited such behavior from scroll and made it more evident with better graphic computing power. Such coherence over products successfully offered seamless and unified "Same UX" to Apple's digital music customers and built strong brand identity.

In summary, Apple wisely took the "Different UI" with its hardware based on marketing, manufacturing cost, and many other factors, yet successfully realized "Same UX" by providing same experience from mechanical wheel to touch-screen flick. As a result, it created a strong user experience identity that contributed to iPod's market dominance.

Fig. 5. Apple iPod (www.apple.com)

5 Speculated Challenges in Deploying "Different UI, Same UX"

5.1 Invisible Battle over UX Governance

In today's IT market, there are three major parties pursuing UX leadership: 1) platform owners (Apple, Google), 2) service providers (Facebook, BBC, AT&T) and 3) manufacturers (Samsung, LG, Nokia). As this market moves rapidly into multi-services with multi-screens and complex ecosystems, these companies' desires to take initiative and drive user experience for their business interest will only intensify. This may create conflicts of interests and result in so-called "battle over UX governance."

Apple and Google, taking advantage of their positions as platform owners, seem to enjoy the highest dominance in UX at the coherence level. But it is also true that many service providers and manufacturers have their own strong consistency policies on their products and services, mostly in the form of UI/UX standardization, and want to extend them to coherence level.

What this implies for "Different UI, Same UX" is that companies would force opponents to take a role in "Different UI" so that they can go with "Same UX." In other words, companies will fight for dominance in the realm of "Same UX." This is a real-world problem for UX designers because it is practically impossible to satisfy all parties' "Different UI, Same UX" needs. The case of Netflix UI in 4.1 was possible because Netflix did not take a strong stance at consistency level and proceeded with minimal elements to secure "Same UX." While this was a wise design decision for all parties involved in the case of Netflix, not all organizations would agree to such a decision.

"Different UI, Same UX" definitely benefits end-users, particularly when we think that end-users will simply view "a set of services enabled by a consortium of partners" as just one service. Each partner in the consortium, however, may not feel the same. Finding resolutions for conflicting desires over UX governance and overcoming the challenge of "who grabs Same UX and who ends up in Different UI" among involved parties will most likely become UX designers' homework. It is almost certain that failing to find resolution and therefore ending up in UX that provides different categories of consistency and coherence at random will confuse users with poor usability. It is thus best to avoid such battle over initiative in UX.

5.2 Branding and “Different UI, Same UX”

One of the core objectives of “Different UI, Same UX” is to achieve high-level UX standardization that contributes branding. It will be UX designers’ responsibility to decide what UI elements to consistently secure across devices and what to intentionally differentiate.

The question remains whether their decisions will be supported by brand managers. There is a possibility brand managers will oppose intentionally differentiating UI. “Different UI” does generate the impression that it goes against “simple and constant sameness,” which is an easy approach to creating a brand. This may even be questioned by other divisions that UX designers are now denying themselves: if consistency is not the best way for branding then what was all that fuss about investing time and effort on UI standardization? It is true many organizations have positioned consistency onto the highest priority in their UI standardization projects.

UX designers will therefore need to convince key decision-makers in branding that “Same UX” via coherence has stronger impact on branding in the long-term sense, and sometimes it could be the only option for UX to contribute to branding. For instance, UX designers in telecommunication organizations should convince brand mangers that “Same UI” is no longer achievable due to heterogeneous nature of their 3-Screen services - particularly considering the fact they don’t make devices but buy them from manufacturers - and that a new branding strategy in UX should be “Different UI, Same UX” to both offer usability satisfaction and deliver one brand identity.

6 Future Work

This study was subsidiary research of a UX for 3-Screens service. Therefore our next step will be a deeper look into an application of “Different UI, same UX” into 3-Screen service. Identifying core UX elements that fall into “Same UX” will be a particularly fruitful challenge.

We believe “Different UI, Same UX” is a design concept not only applicable to 3-Screens but to broader range of design domains. A good example would be UX standardization of various kinds of products manufactured by one organization, for which further research focused on finding relevant industrial cases is needed. Lastly, proposing solutions for the speculated challenge described in chapter 5 will be a long-term mission of this study.

References

1. Anderson, J., McRee, J., Wilson, R.: The Effective UI Team.: Effective UI-The Art of Building Great User Experience in Software. O’Reilly Media, Sebastopol (2010)
2. Nielsen, J.: Coordinating User Interfaces for Consistency, The Morgan Kaufmann Series Interactive Technologies, San Francisco, CA, USA (1989)
3. Dubberly, H., Evenson, S.: The Experience Cycle. Interactions 15(3) (March and April 2008)

4. Roto, V., Rautava, M.: User Experience Elements and Brand Promise. In: International Engagability & Design Conference (Idec4), in Conjunction with NordiCHI 2008 Conference, Lund, Sweden (2008)
5. Pyla, P.S., Tungare, M., Holman, J., Pérez-Quiñones, M.: Continuous User Interfaces for Seamless Task Migration. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5612, pp. 77–85. Springer, Heidelberg (2009)
6. Trimeche, M., Suomela, R., Aaltonen, A., Lorho, G., Dossaji, T., Aarnio, T., Tuoriniemi, S.: Enhancing end-user experience in a multi-device ecosystem. In: MUM 2005 Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia. ACM, New York (2005)
7. Nielsen, J.: Usability Engineering. Academic Press, Boston (1993)
8. Kim, D., Yoon, W.: A Method for Consistent Design of User Interaction with Multifunction Devices. In: HCI International 2009, Human Centered Design, pp. 202–211 (2009)
9. Meskens, J., Luyten, K., Coninx, K.: Jelly: A Multi-Device Design Environment for Managing Consistency Across Devices. In: International Working Conference on Advanced Visual Interfaces (AVI), Rome, Italy (2010)
10. Grigoreanu, Fernandez, R., Inkpen, K., Robertson, G.: What designers want: Needs of interactive Application Designers. In: Proc. VL/HCC 2009, IEEE Computer Society, Los Alamitos (2009)
11. Segerståhl, K.: Crossmedia Systems Constructed around Human Activities: A Field Study andImplication for Design. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 354–367. Springer, Heidelberg (2009)
12. Segerståhl, K.: Utilization of Pervasive IT Compromised? Understanding the Adoption and Use of a Cross Media System. In: Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia, Umeå, Sweden (2008)
13. Segerståhl, K., Oinas-Kukkonen, H.: Distributed user experience in persuasive technology Environments. In: de Kort, Y.A.W., IJsselsteijn, W.A., Midden, C., Eggen, B., Fogg, B.J. (eds.) PERSUASIVE 2007. LNCS, vol. 4744, pp. 80–91. Springer, Heidelberg (2007)
14. Feiner, S.K.: Environment Management for Hybrid User Interfaces. IEEE Personal Communications, 50-53 (October 2000)
15. Filgueiras, L., Correa, D., Oliveira Neto, J., Facis, R.: X-gov planning: how to apply cross media to government services. In: ICDS 2008, pp. 140–145. IEEE, Los Alamitos (2008)
16. Seffah, A., Javahery, H.: Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces. Wiley & Sons, Ltd., Chichester (2005)
17. Wäljas, M., Segerståhl, K., Väänänen-Vainio-Mattila, K., Oinas-Kukkonen, H.: Cross-platform service user experience: a field study and an initial framework. In: MobileHCI 2010 - Human Computer Interaction with Mobile Devices and Services, Lisbon, Portugal (2010)

Measurement of User Experience to Select a Comfortable Mattress

Jung-Yong Kim[1], Seung-Nam Min[1], Min-Ho Lee[1], Joo-Hyun Jeong[1], Jung-Ho An[2], and Young-Sung Shin[2]

[1] Department of Industrial & Management Engineering, Hanyang University, Ansan, Republic of Korea
jungkim@hanyang.ac.kr
[2] Research Laboratory, Simmons Co. Ltd, Gyeonggi-do, Republic of Korea
cool821@simmons.co.kr

Abstract. This study was designed to develop a methodology measuring the user experience with mattress both in the past and showroom, and to eventually recommend a healthy and comfortable mattress for individual user. Five mattresses with different hardness were used to find the most compatible mattress with individual subject's physical and psychological condition. User experience such as lying on the mattress in showroom was analyzed by quantitatively measuring Electromyography of low back muscle, heart rate change, and oxygen saturation level. In addition, the whole body pressure distribution was measured to examine the dermal discomfort. A questionnaire was used to record the past personal experience and preference on mattress. A selection rule with the finally chosen four independent variables and mathematical scale was developed to find the best mattress for individual. Furthermore, a regression analysis was performed to predict the level of muscle relaxation in order to have the least measuring process in the showroom. The Body-Mattress Compatibility Score (BMCS) indicating the proper level of hardness was computed in this study and compared with subjective satisfaction score for validation, and it was found that ten out of twenty subjects showed the same score, and other ten subjects showed only one score difference.

Keywords: user experience, mattress selection, physical and psychological comfort, pressure distribution, body mattress compatibility, showroom.

1 Introduction

A sleep is one of the most important parts of daily life, we all know that the quantity and quality of sleep are greatly associated with the release of mental stress as well as physical fatigue. Therefore, a sufficient sleep and comfortable bed are very essential to maintain healthy physical and mental conditions [1].

In Korea, there are an increasing number of people using western style bed. However, they are often fastidious in choosing the right mattress for them. In fact, people use their past experience with their old mattress as well as the spontaneous experience they encounter in showroom to finally decide to buy a bed. People often lie down on

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 449–458, 2011.

the bed to examine how comfortable they are, and each individual makes a decision by using their own intuitive judgment. However, there is no warranty whether or not such decision could reflect their experience objectively and wisely.

Regarding the hardness of mattress, Suckling et al., (1957) reported that a hard mattress disturbed a sound sleep and made people toss and turn. Parson (1972) found that appropriate hardness of bed mattress was good for a sound sleep. Moreover, the hardness of mattress is associated with reducing the muscle pain [4].

Yu et al., (2009) used the level of physical relaxation of individual subject to evaluate different mattresses by monitoring psycho-physiological biofeedback. Yu (2010) reported that a biomechanical support from the hard mattress can make low back muscle relax. Lahm and Jaizzo (2002) observed that mattresses not supporting the upright spine did increase discomfort in the back and waist muscles.

Kim et al., (2007) reported that subjective comfort increased when skin pressure was equally distributed across mattress surface and skin temperature rise was minimal. Pressure concentration on a particular body part could cause a skin problem as well as increasing subjective discomfort [8]. The pressure distribution was described as one of the most important variables in designing comfortable mattress [2]. Also, it was reported that the pressure should be evenly distributed across major body parts (head, torso, waist and legs) so that the spine would not be overly bent or curved [4].

Therefore, the purpose of the study is to develop a methodology to measure user experience with mattress in the past and in the showroom as objectively as possible by using the multiple variables used in previous studies, in order to recommend a healthy and comfortable mattress for individual customer.

2 Methodology

In this study, it was a basic assumption that we would be able to find the most comfortable and healthy mattress for individual customers if we knew their previous experience as well as their physical and psychological experience in the showroom, possibly in a very quantitative manner (Fig. 1).

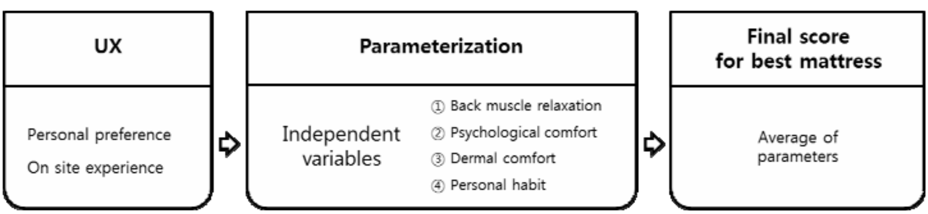


Fig. 1. Conceptual summary of methodology to measure the user experience with mattress

It is no need to argue that a good mattress can provide a good feeling of physical and psychological relaxation of the body when lying on it. Therefore, in this study, the back muscle relaxation, psychological comfort, cardiac efficiency, dermal comfort, and personal habit were measured by EMG, heart rate change, body pressure concentration, oxygen saturation level, and personal questionnaire. It is reasonable to say that those individual variables used in this study would evenly reflect the level of overall comfort. In order to express that, Body-Mattress Compatibility Score (BMCS) was suggested in this study as the average value of normalized score of individual variables. In addition to that, a regression analysis was performed to estimate the EMG value, which is difficult to be measured in the showroom. The final result with the estimated value was compared with subjective satisfaction score for validation. The Figure below shows the study flow (Fig. 2).

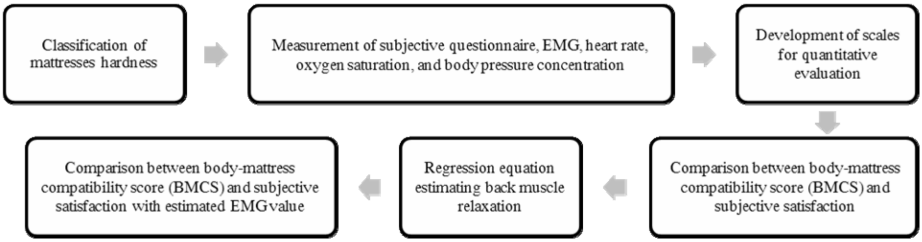


Fig. 2. Study flow-chart

3 Classification of Mattress Hardness

Five mattresses built with different hardness produced by “S” company were used for the present study. The hardness of the mattresses were measured by an endurance tester (KS-200B, Korea) by pushing the mattress material as much as 80 mm. The difference in hardness was significant among five mattresses and showed in Table 1. The hardness was measured five times and five different spots per mattress.

Table 1. Result of paired t-test on mattresses with different hardness

	mattress 1	mattress 2	mattress 3	mattress 4	mattress 5
mattress 1	-	.000***	.000***	.000***	.000***
mattress 2	.000***	-	.000***	.000***	.002***
mattress 3	.000***	.000***	-	.000***	.000***
mattress 4	.000***	.000***	.000***	-	.000***
mattress 5	.000***	.002***	.000***	.000***	-

***p<0.01

Depending on the result, we could confirm that the five mattresses could be classified into five levels based on the hardness. We defined that score 5 is the hardest and score 1 is the softest one.

4 Measurement of Subjective Questionnaire, EMG, Heart Rate, Oxygen Saturation, and Body Pressure Concentration

4.1 Subjects

An experiment was designed to measure the various bio-signals from ten male and ten female subjects who were between 20 to 30 years old (age 27.1±6.4, height 167.1±7.2, weight 66.2±18.4). None of them had a medical history related to back pain (Table 2).

Table 2. Information on participants

	Age	Height(cm)	Weight(kg)
Average(SD)	27.1(6.4)	167.1(7.2)	66.2(18.4)

4.2 Experimental Design

Five independent variables were the back muscle relaxation (EMG), psychological comfort (heart rate), dermal comfort (body pressure concentration), personal habit (personal questionnaire), and cardiac efficiency (oxygen saturation) (Table 3).

Table 3. Independent variables used to measure the comfort of the mattress

Measured variables	Definition	Formula
Back muscle relaxation	Decrement of muscle activity before and after using a mattress	$\text{Back muscle relaxation}_{\text{i}} = \frac{\text{IEMG}_{i(lying\,50s\sim60s)}}{\text{IEMG}_{i(standing\,50s\sim60s)}} (i = 1,2,...,20)$
Psychological comfort	Decrement of heart rate before and after using a mattress	$\text{Psychological comfort}_{\text{i}} = \frac{\text{HR}_{i(lying\,50s\sim60s)}}{\text{HR}_{i(standing\,50s\sim60s)}} (i = 1,2,...,20)$
Dermal comfort	The ratio of peak pressure to mean pressure when lying on a mattress	$\text{Dermal comfort}_{\text{i}} = \frac{\text{P}_{average\,i}}{(\text{P}_{1st\max i} + \text{P}_{2nd\max i} + \text{P}_{3rd\max i})/3} (i = 1,2,...,20)$
Personal habit	Likert score from personal questionnaire	$\text{Personal habit}_{\text{i}} = \frac{\text{Q1}_{\text{i}} + \text{Q2}_{\text{i}} + \text{Q3}_{\text{i}} + \text{Q4}_{\text{i}}}{4} (i = 1,2,...,20).$
Cardiac efficiency	Increment of oxygen saturation in peripheral vessels before and after using a mattress	$\text{Cardiac efficiency}_{\text{i}} = \frac{\text{SaO}_{2i(lying\,50s\sim60s)}}{\text{SaO}_{2i(standing\,50s\sim60s)}} (i = 1,2,...,20).$

4.3 Apparatus

To measure back muscle relaxation, two EMG electrodes out of eight channel EMG system (ME-6000T, Mega Win, Kupio, Finland) were used to both left and right side of the erector spine. A pad (190cm x 80cm) with 32 x 32 digital film sensors was used to measure the level of concentration of body pressure. Nihon Kohden Company's pulse oximeter was used to measure the heart rate and the oxygen saturation.

The questionnaire included 5 point scale questions about personal information, history of back pain, the preference on hardness of mattress in the past.

4.4 Procedure

Before the experiment, anthropometric data of subjects were measured, and a questionnaire was filled out. Then, the experiment process was explained and measuring device was attached to subjects and the stability of the signal was checked. EMG, heart rate and oxygen saturation signals were measured for 60 seconds while subjects were standing (baseline data) and lying (relaxation data) on the mattress. The data collected between 50 to 60 seconds, which was the most stable data, were used for data analysis. Also, the body pressure concentration was measured while subjects were lying on the mattress. Upon finishing the measurement, subjects were asked to lie down on other five mattresses one by one, and they recorded the level of subjective satisfaction in five Likert scale.

5 Development of Scales for Quantitative Evaluation

The normality of data of all variables was examined (Table 4). As results, the oxygen saturation (SaO_2) was excluded in evaluation process since it did not show a normal distribution.

Table 4. One-sample kolmogorov-smirnov for examination of normal distribution

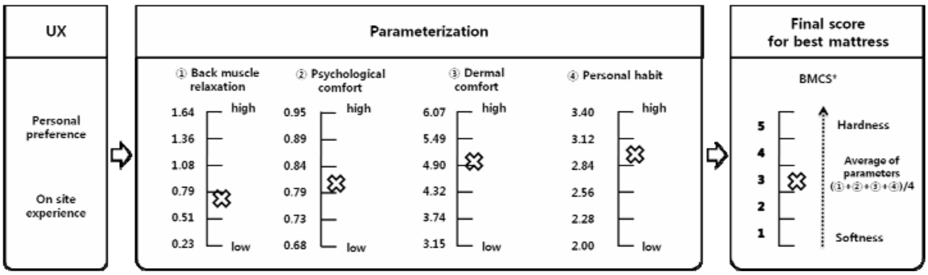
	Back muscle relaxation	Psychological comfort	Dermal comfort	Personal habit	Cardiac efficiency
N	20	20	20	20	20
Kolmogorov-Smimov's Z	0.694	0.722	0.422	0.909	1.790
Approximate significance probability(two-sided)	0.722	0.674	0.994	0.381	0.003***
Normal distribution(Yes/No)	Yes	Yes	Yes	Yes	No

***p<0.01, **p<0.05, *p<0.1,

Table 5. Final scores of independent variables (raw score and normalized score in 5 point scale)

	Back muscle relaxation		Psychological comfort		Dermal comfort		Personal habit	
	Raw score	Nor. score	Raw score	Nor. score	Raw score	Nor. score	Raw score	Nor. score
Sub1	1.24	2	0.68	5	4.50	3	3.0	3
Sub2	0.99	3	0.81	3	5.47	2	2.0	2
Sub3	0.70	4	0.80	3	5.91	1	2.8	3
Sub4	0.54	4	0.78	3	4.17	4	2.8	3
Sub5	0.70	4	0.75	4	5.08	2	2.8	3
Sub6	0.86	3	0.89	1	3.89	4	2.4	2
Sub7	0.67	4	0.95	1	4.16	4	3.2	3
Sub8	0.55	4	0.85	2	4.80	3	3.4	3
Sub9	0.47	5	0.81	3	3.45	5	3.2	3
Sub10	0.66	4	0.90	1	5.32	2	3.4	3
Sub11	0.56	4	0.72	5	4.73	3	3.0	3
Sub12	1.07	3	0.72	5	6.07	1	3.4	3
Sub13	0.40	5	0.87	2	3.15	5	2.4	2
Sub14	0.76	4	0.78	4	5.45	4	3.2	3
Sub15	0.74	4	0.88	2	3.99	4	2.8	3
Sub16	1.16	2	0.83	3	4.97	2	3.0	3
Sub17	1.64	1	0.80	3	5.90	1	2.6	3
Sub18	1.36	2	0.73	4	4.16	4	3.4	3
Sub19	0.87	3	0.82	3	3.81	4	3.2	3
Sub20	0.23	5	0.88	2	4.96	2	2.2	2

Final variables used to develop a scale were the back muscle relaxation, psychological comfort, dermal comfort, and personal habit as shown in Table 5. Based on the maximum and minimum score of individual variable, 1 to 5 scale was used in the study. The Interval of the scale was determined as {(maximum – minimum)/5}. Score 5 indicates that subject can relax on the mattress very well. Then, the hardest mattress can be recommended for biomechanical advantage to support low-back area. However, even the hardest mattress, which is currently in the market, has a substantial cushion on the top of it to make user feel comfortable (fig. 3). If the score is close to 1, even softer mattress will be recommended for user feel more relaxed.



*BMCS= (normalized score of psychological comfort + normalized score of dermal comfort + normalized score of back muscle relaxation + normalized score of personal habit) / 4

Fig. 3. The scores and scales of individual variables, and body-mattress compatibility score (BMCS)

6 Comparison between Body-mattress Compatibility Score (BMCS) and Subjective Satisfaction

For validation of the BMCS, the score from subjective satisfaction was used. The comparison result is in Figure 4.

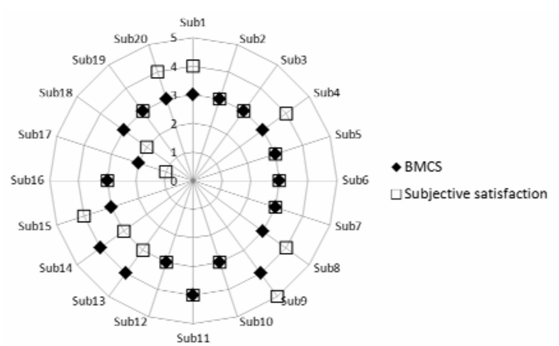


Fig. 4. Comparison between BMCS and subjective satisfaction scales

BMCS and subjective satisfaction were matched each other among ten subjects. Among another ten subjects, BMCS and subjective satisfaction scales were found to be different only by one point.

7 Regression Equation Estimating Back Muscle Relaxation

Since it was difficult to use EMG equipment in showroom, an alternative method was investigated. By observing the spearman correlation coefficients table, the back muscle relaxation determined by measured EMG had meaningful relationship with other variables. Thus, a multiple regression analysis was conducted to investigate the possibility to estimate EMG value instead of actually measuring it.

Table 6. Correlation analysis with each variable

	Back muscle relaxation	Psychological comfort	Dermal comfort	Personal habit
Back muscle relaxation	1.000	0.580*	0.581*	0.520*
Psychological comfort	0.580*	1.000	0.653**	0.349
Dermal comfort	0.581*	0.653**	1.000	0.260
Personal habit	0.520*	0.349	0.260	1.000
Oxygen saturation	0.454	0.110	0.170	0.377

***p<0.01, **p<0.05, *p<0.1,

In the regression model (equation 1) using three independent variables such as dermal comfort, psychological comfort and personal habit indicated R^2 of 0.512.

$$Y = 2.077- 0.003X_1 + 0.504X_2 - 0.178X_3. \tag{1}$$

Y: Back muscle relaxation
\X_1: Psychological comfort
X_2: Dermal comfort
X_3: Personal habit.

8 Comparison between Body-mattress Compatibility Score (BMCS) and Subjective Satisfaction with Estimated EMG Value

BMCS and subjective satisfaction were matched each other among ten subjects. Among another eight subjects, BMCS and subjective satisfaction scores were found to be different only by one point. Two subjects showed two points difference.

9 Discussion

The study has attempted to measure user experience in a very quantitative manner unlike many UX methods using mostly intuitive logics, although this study has intuitively adopted the idea that physical and psychological relaxation would be an important factor determining overall comfort of the mattress in the ideation process. We may say that user experience itself has multi-dimensional attributes that we cannot simply measure only through an experimental observation. However, it can be challenged when the behavioral pattern of the experience is simple and repetitive enough; and we have equipments and techniques to measure and analyze such human behavior.

In showroom situation, people buy products by using their own judgment based on their past and on site experience. However, we do not know whether or not buyers make the best decisions at all. Thus, it would be very reassuring for buyers if we could provide a methodology to help them guide or confirm their decisions by showing statistical results in ten minutes after interesting science experiment. Therefore, the methodology in this study can be considered as a useful tool for both merchandisers and customers in buying mattress.

In fact, we used four independent variables to cover various physical and psychological experiences of subjects, which induced the risk of mathematical fuzziness during the process of handling multiple variables with different dimensions. The measured values were transformed into normalized scores to have a unified value summarizing the results from the four final variables used in this study. In that process, it was also assumed that each variable had the same importance or same weight to determine the final result. In reality, this assumption may not be correct since people usually have their own preference or bias in reaching to their own final decision. So, this can create the insensitivity problem of the methodology if somebody's buying decision is made by one particular factor rather than the average of many different factors. This issue needs to be tackled in future study as we accumulate the field data to improve the sensitivity of the methodology.

For now, in order to confirm the validity of BMCS result, a subjective satisfaction was used in this study. Interestingly, there was no big difference between BMCS and subjective satisfaction score in selecting the final mattress. This validation process allowed us to use the methodology in the field without too much worry about producing conflicting outcome against buyer's preference. Statistically speaking, people tend to choose the mattress with medium hardness rather than extreme one. The methodology also guides customers to choose mattress based on the feedback from various psycho-physiological sensors, of which values were normally distributed. That is, we can expect that BMCS, that is the average of four scores from individual variables, is normally distributed and well depicts natural human behavior.

Another hidden device to prevent unreasonable outcome is in the questionnaire asking personal habit and previous experience about the mattress. There were two questions with the same content only with different wording. If a customer give different answers for the same questions, all the data from questionnaire is discarded to prevent biased outcome. This is probably the case when customer is not attentive or sincere enough in answering the question or not understanding the question. This process also helps us not having unreasonable recommendation.

The methodology introduced in this study was developed as a commercial system used in actual showroom. All the interfaces for easy operation of measuring process was designed and integrated as one complete system (Fig. 5). Visiting customers take approximately ten minutes to go through the test to get the final printed results. They experience an interesting science experiment which is very new and curious event for them, which actually help them to make a decision in buying the right mattress. In future, the validity and sensitivity of the methodology can be observed and analyzed based on empirical data collected from the field. Furthermore, a better mathematical approach can be investigated to improve the efficacy of the system.

Fig. 5. The system set up in showroom and an example of printed result

10 Conclusion

We have shown that the user experience could be quantitatively measured for practical use in actual showroom. The methodology used in this study was designed for a particular product, but it could be applied to other products with a minor revision in various situations. The validity and sensitivity of the methodology can be further improved as we accumulate the field data and employ a more sophisticated device to integrate multi-dimensional information of human behavior in the near future.

References

1. Donaldson, E., Kennaway, D.J.: Effects of Temazepam on Sleep, Performance, and Rhythmic 6-sulphatoxymelatonin and Cortisol Excretion after Transmeridian Travel. Aviat Space Environ. Med. 62(7), 654–660 (1991)
2. Gaby, G.B., Sten, E.: The Influence of Bed Hardness on Sleep Quality. Applied Ergonomics 31, 487–497 (2000)
3. Kim, J.Y., Park, J.S., Yoo, S.W., Lee, S.J., Goh, T.S.: Ergonomic Test of Air-ball Mattresses Using Skin Pressure and Temperature Change. Proceeding of the Ergonomics Society of Korea 2(1), 283–287 (2007)
4. Kovacs, F.M., Abraira, V., Pena, A., Martin-Rodriguez, J.G., Sanchez-Vera, M., Ferrer, E., Ruano, D., Guillen, P., Gestoso, M., Muriel, A., Zamora, J., Gil del Real, M.T., Mufraggi, N.: Effect of Hardness of Mattress on Chronic Non-specific Low-back Pain: Randomised, Double-blind, Controlled, Multicentre Trial. Lancet 362, 1599–1604 (2003)
5. Lahm, R., Iaizzo, P.A.: Physiologic Responses during Rest on a Sleep System at Varied Degrees of Hardness in a Normal Population. Ergonomics 45(11), 798–815 (2002)
6. Parsons, H.M.: The Bedroom. Human Factors 14(5), 42–450 (1972)
7. Suckling, E.E., Koenig, E.H., Hoffman, B.F., Brooks, C.: The Physiological Effects of Sleeping on Hard or Soft Beds. Human Biology 29, 274–288 (1957)

8. Vanderwee, K., Grypdonck, M., Defloor, T.: Alternating Pressure Air Mattresses as Prevention for Pressure Ulcers: A literature review. Age and Ageing 34(3), 261–267 (2005)
9. Yu, S.W., Kim, J.Y., Min, S.N., Sung, S.H.: Analysis of Suitability for Mattresses by Using Psycho-Physiological Measures. Proceeding of the Korean Society for Emotion & Sensibility 19, 63–66 (2009)
10. Yu, S.W.: Analysis of Suitability for Mattresses by Using the Level of Physical Relaxation. master's thesis (2010)

EMA: Automated Eye-Movement-Driven Approach for Identification of Usability Issues

Oleg V. Komogortsev, Dan E. Tamir, Carl J. Mueller,
Jose Camou, and Corey Holland

Department of Computer Science, Texas State University-San Marcos,
San Marcos, TX 78666, USA
{ok11,dt19,cm58,jc1782,ch1570}@txstate.edu

Abstract. The work described in this paper presents an automated, eye movement-driven approach (EMA) that allows for the identification of time intervals in which a user is experiencing difficulties in locating interface components required for completion of a task. Due to the substantial amount of visual search exhibited during these time intervals, this type of the user behavior is referred to as excessive visual search (ES). In this work we propose and evaluate several ES detection algorithms as part of the EMA. Empirical results indicate that it is possible to identify ES with a certain degree of accuracy (51-61% on average), warranting future research that would allow for increased accuracy in ES identification and reduction of misclassification errors. Practical application of EMA should allow the reduction of the amount of time required for manual detection of usability problems present in graphical user interfaces.

1 Introduction

Visual assessment of a user's behavior through the analysis of screen recordings is extensively employed in usability studies to identify issues with interface design and implementation [1, 2, 6, 11]. Such assessment is frequently time consuming and requires a trained human observer capable of identifying time intervals in which a user is experiencing difficulties completing task requirements. Such difficulties usually include prolonged searches for interface components, navigation issues, data entry difficulties, etc. As a result, visual assessment of a video recording presents the following challenges: 1) it cannot provide useful information when a user does not employ the mouse or keyboard while searching for interface components; 2) the length of the recording might be extensive, and a substantial amount of time may be required in identifying the intervals in which a user is exhibiting *excessive visual search* (ES) to locate necessary interface components.

In this paper we attempt to resolve these challenges through the automated analysis of the *Human Visual System* (HVS) behavior represented by a sequence of the various basic eye movements and their properties. Assessment of HVS behavior is possible with the aid of an eye tracking device that records raw eye-gaze data [3], and eye movement classification algorithms that translate this data into a sequence of basic eye movements of varying type [8]. Following study and experimentation, we propose

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 459–468, 2011.

an automated, eye movement-driven approach (EMA) that allows for the identification of ES intervals in a recording of a user accomplishing a task, thereby reducing the amount of time necessary for manual inspection of usability problems.

Challenges associated with the EMA's design and evaluation include: the definition of a baseline that can be employed in validating the accuracy of ES classification algorithms, meaningful automated segmentation of user recordings, and the identification of the segments that relate to ES. Present work discusses possible solutions to these challenges and presents results from a case study in which EMA was employed.

2 Theory

2.1 User Behaviors

When a user has to accomplish a task, they might exhibit the following behaviors: *Task completion* (TC) - performing the operations necessary to complete a given task; *Excessive visual search* (ES) – experiencing prolonged difficulties finding the interface components necessary for task completion. TC might contain moderate search behavior where a user is moving from one component necessary for task completion to the next without spending an excessive amount of time in locating the required components. Both TC and ES depend on the user, application, and instructions. Depending on an application and the type of instructions, two additional user behaviors can be exhibited: *Off-screen* (OS) - reading task-related instructions presented outside the boundaries of the computer monitor; *Idle* (IL) – waiting for interface response after a specific action is performed.

2.2 Human Visual System

Given an interface which is void of smooth moving objects, the HVS primarily exhibits two eye movement types: fixations, dwelling on a stationary target of interest that results in high acuity visual information supplied to the brain; and saccades, extremely rapid eye rotations that transition the eye globe between points of fixation [3].

Because a user employs fixations to collect visual information, a large amount of fixations can represent search difficulties [5]. Long fixation durations can indicate difficulties in extracting the information required for task completion [7]. Large amounts of saccades might indicate increased searching [5], as the number of saccades is directly proportionate to the number of exhibited fixations. Large saccade amplitudes might be indicative of large distances between the components that are required for task completion. In addition, the duration of a scanpath (sequence of fixations and saccades) might represent non-efficient searching when it becomes too long [4]. Moreover, eye tracking equipment enables the extraction of information about pupil diameter. Pupil diameter might be indicative of the user's cognitive workload [9], possibly correlating with a case in which the user tries to find information required for task completion.

3 Implementation

The following section presents considerations that were employed in producing the classification baseline achieved via manual classification of user recordings. Proper construction of a baseline is very important because it can be treated as a “golden standard” which can be compared to automated ES classification methods for accuracy verification. This description is followed by the logic used for automated segmentation of user recordings, and the description of algorithms employed for automated classification of excessive search behavior.

3.1 Manual Classification of User’s Behaviors

A trained research assistant performed manual classification of the task recordings with superimposed eye movement traces, in order to build a classification baseline. Among the four behaviors described in Section 2.1, each type has its own unique implications and criteria: ES and OS behaviors were identified primarily by eye movement data; TC behavior was identified primarily by mouse/keyboard events; and IL behavior by consideration of the application state.

ES was defined as any onscreen search interval not directly related to TC. Such behavior occurs when a user has difficulty locating, or manipulating, the target of their current task, and can indicate problems in the usability of an application. Intervals of the recording with erratic scanpaths, unduly long fixation durations, or frequent shifts in regional attention are marked as ES.

TC behavior is defined as any onscreen search interval directly contributing to the completion of the user’s current task, typically identified by mouse/keyboard input and the immediately preceding scanpath(s). Such behavior occurs when the user has located the target of their current task. A well designed application would elicit only TC behavior. Intervals of the recording from when the user located the target of their current task until manipulation of the target element ceased are marked as TC behavior.

IL behavior is defined as any user actions taken while the user is unable to interact with the application, identification of which requires knowledge of the application’s state. While ES can indicate problems in the usability of an interface, IL behavior can indicate problems in the implementation of the underlying application logic. Intervals of the recording during which interface components are being loaded, file input/output is being performed, or the interface is otherwise unresponsive are marked as potential candidates for IL behavior.

OS behavior is defined as any interval in which the user’s gaze is directed off-screen, and indicates a shift in the user’s attention away from the current task. Intervals of OS behavior are identified using the scanpath overlay of the usability recording, further analysis being unnecessary.

While these behavioral states can be discretely defined and identified, they are not mutually exclusive. That is, there is a certain amount of overlap between these states. Examples include: typing data into a text box while reading the value off-screen (TC / OS); staring for an extended duration at a small interface component while positioning the mouse for selection (ES / TC); scanning the screen while waiting for a progress bar to indicate that file output has completed (ES / IL). As a result, an order of precedence for categorizing an interval that shared the characteristics of multiple behaviors is required.

ES behavior, identifying tangible problems in interface usability, is given the highest precedence, followed in order by IL behavior, TC behavior, and, finally, OS behavior. That is, a candidate interval for both IL and OS behavior is considered only as IL behavior, and a candidate interval for both ES and TC behavior is considered only as ES. Based on this precedence, the timeline for each usability recording is completely segmented into these four behaviors, with the beginning of each behavior interval extending to the beginning of the next behavior interval.

3.2 Automated Classification of Excessive Search

Among user behaviors, ES most directly relates to usability issues; therefore, the present implementation of EMA targets only automated detection of ES. To allow for the automated detection of ES intervals, the EMA consists of two parts: segmentation of task recording and identification of excessive search intervals.

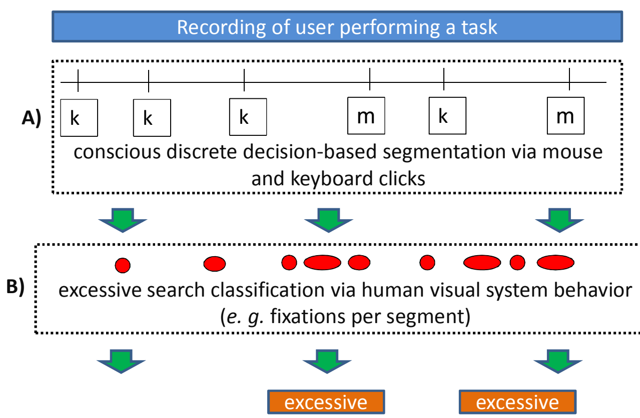


Fig 1. Excessive search classification by automated eye-movement-driven approach. A) Segmentation. B) Classification.

Segmentation of Task Recording. The goal of a segmentation method is to identify time intervals that can be subsequently mapped to user behaviors. It can be hypothesized that eye tracking data is capable of providing the information required to detect subtle transitions in user behavior; however, to the best of our knowledge, no such methods exist at this time. The method proposed in this work performs a segmentation based on the user's conscious interaction decisions, represented by mouse and keyboard events. Specifically, when the user is performing a task, each subsequent mouse or keyboard event ends one behavior segment and begins the next (Figure 1.A).

Identification of Excessive Search Intervals. The following threshold-based methods were employed for the automated classification of ES among segmented behavior intervals.

Fixation Based Algorithm
The fixation based algorithm (ES-F) employs the fixation count of each interval as its index value, and the average fixation count as the threshold. Behavior intervals with a fixation count above the threshold are classified as ES. Figure 1.B presents an example.

Saccade Based Algorithm
The saccade based algorithm (ES-S) employs the average saccade amplitude of each interval as its index value, and the average of average saccade amplitudes as the threshold. Behavior intervals with average saccade amplitude above the threshold are classified as ES.

Pupil Based Algorithm
The pupil based algorithm (ES-P) employs the average pupil dilation of each interval as its index value, and the average of average pupil dilations as the threshold. Behavior intervals with average pupil dilation below the threshold are classified as ES.

Scanpath Length Algorithm
The scanpath length algorithm (ES-SL) employs the total saccade amplitude of each interval as its index value, and the average across all intervals as the threshold. Behavior intervals with saccade amplitude above the threshold are classified as ES.

4 Experiment Setup

4.1 Software

Usability testing was performed on two versions (v9.3 and v10.3) of Emerson Process Management's DeltaV Control Studio as part of a related but separate study by researchers at Texas State University for developing methods of objective usability evaluation. Screen recordings and their corresponding scanpath/input event overlays, shown in Figure 2, were viewed for manual classification of user behaviors with Tobii Studio [12] software. All algorithms for the automated data analysis were implemented and executed in MATLAB. A velocity threshold algorithm (I-VT) was used to reduce the eye movement data into the fixations and saccades [8].

4.2 Apparatus

Usability testing was conducted with the Tobii X120 eye tracker [12] running at 120Hz. DeltaV was run on a Dell Optiplex 745 with 4 GB of RAM.

4.3 Participants

A total of 14 student volunteers and 12 expert users participated in the usability testing. Recordings were used from only 10 of these in the current context (selected randomly), due to the substantial amount of time required for manual classification.

4.4 Procedure

Participants were asked to complete a series of 15 tasks with DeltaV, during which screen recordings, eye movement records, and input logs were generated and synchronized for each task. Tasks were similar to each other with enough variation to reduce learning effects. The details of task preparation are described elsewhere [10].

ES intervals were manually classified for 10 of the recordings (chosen arbitrarily from unique subjects with a uniform distribution of trials). The EMA was employed to extract the ES interval from the same recordings. Results of both classifications were compared to assess the accuracy of the automated classification algorithms.

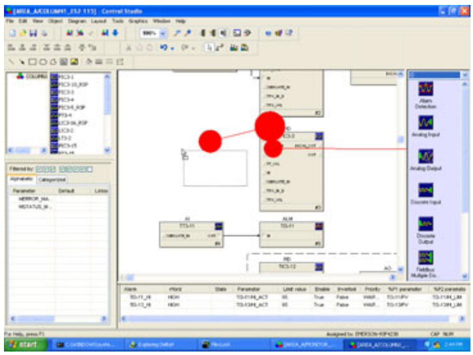


Fig. 2. DeltaV v9.3 Control Studio

4.5 Performance Metrics

The following metrics are employed to determine the accuracy of the ES classification algorithms.

Average Percent of Total Time Classified. The average percent of the total time classified as ES by a given algorithm is:

$$Percent\_of\_Total = 100 \bullet \frac{total\_classified\_time}{total\_recording\_time} \quad (1)$$

Average Percent Correctly Classified. The percent correct is defined as the average percent of time correctly identified as ES when compared to the time intervals produced by manual classification.

$$Percent\_Correct = 100 \bullet \frac{correctly\_classified\_time}{total\_manual\_time} \quad (2)$$

Average Percent Error. The percent error is defined as a combination of the average percent of time misclassified as ES and the average percent of manually classified time that was not classified by the algorithm, as compared to the time intervals produced by manual classification.

$$Percent\_Error = 100 \bullet \frac{miscassified\_time + unclassified\_time}{total\_manual\_time} \quad (3)$$

5 Results

5.1 Manual Classification of Users' Behaviors

Figure 3 presents results. The amount of classified TC behavior was substantial and close to 40%. The amount of ES was approximately 16%, suggesting that, in the case of an accurate algorithm, it would be necessary to review only 16% of the recording in order to identify usability problems related to the placement of interface components required for task completion. It is important to note that the amount of time spent by users in ES behavior was approximately one third that of TC behavior, indicating that a substantial amount time can be potentially saved as a result of interface optimization.

It is possible to notice that the amount of OS behavior is large (approximately 40%) because users were frequently reading the task instructions presented off-screen. The amount of IL behavior was small (close to 4%). The difference in time between different behaviors was statistically significant, $F(3, 36) = 13.6, p < 0.001$.

5.2 Automated Classification of Excessive Search

Average Percent of Total Time Classified. Figure 4 presents results. Among the considered algorithms, ES-F on average classified the lowest percent of the total recording time (34%) as excessive search. The other three methods yielded similar average times that were close to 50% with varying amounts of deviation. A one-way ANOVA test indicated no significant main effect between algorithms, $F(3, 36) = 1.49$, $p = 0.23$.

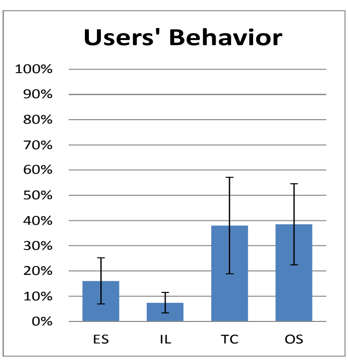


Fig. 3. Manually classified users' behaviors out of total recording time

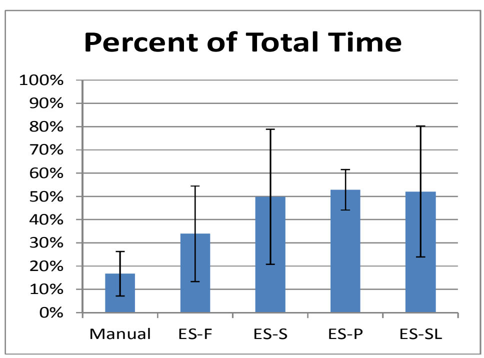


Fig. 4. Excessive search: average percent of total time classified

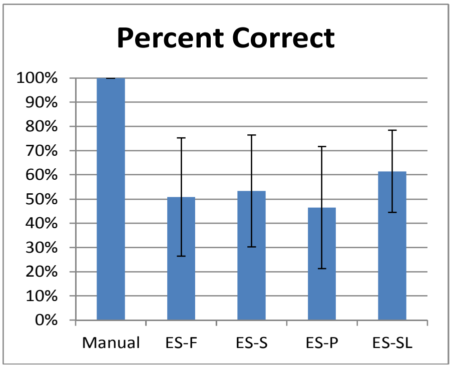


Fig. 5. Excessive search: average percent correct

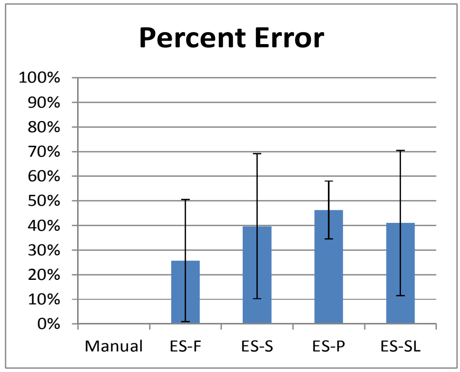


Fig. 6. Excessive search: average percent error

Average Percent Correctly Classified. Figure 5 presents results. ES-SL had the highest average percent of correctly identified intervals (61%), while ES-P had the lowest percent (46%). ES-F and ES-S had similar performance that was close to 52%. A one-way ANOVA test indicated no significant main effect between algorithms, $F(3, 36) = 0.77$, $p = 0.52$.

Average Percent Error. Figure 6 presents results. Among the considered algorithms, ES-F had the lowest average percent error (26%), while ES-P had the highest error (46%). The ES-S and ES-SL algorithms had similar performance that was close to 40%. A one-way ANOVA test indicated no significant main effect between algorithms, $F(3, 36) = 1.24$, $p = 0.31$.

6 Discussion

6.1 Best Algorithm for Excessive Search Classification

Among the ES algorithms considered in this work, the fixation-based algorithm (ES-F) can be selected as the performer with the least error and the third best classification performance. The scanpath-based algorithm (ES-SL) yielded the highest percent of correct classification with the third smallest error. Therefore, ES-F can be considered as a more conservative algorithm in terms of classification error, while ES-SL provides the highest classification accuracy among considered algorithms. The actual choice between ES classification algorithms should be made according to the requirements of the usability experiment.

6.2 Worst Algorithm for Excessive Search Classification

Among the ES algorithms considered in this work the pupil dilation-based algorithm performed most poorly, providing the lowest percent of correct classification and producing the largest misclassification error. Therefore, it is possible to hypothesize that excessive searching does not correlate well with increased pupil diameter. Additional research is necessary to prove or disprove this hypothesis.

6.3 Limitations

Manual Classification. Regardless of the instructions and considerations presented in Section 3.1, manual classification of user behaviors is a subjective process. One of the possible solutions to this issue is a baseline created by considering several manual classifications performed by different people.

Automated Classification. The current record segmentation method only considers mouse and keyboard events as boundaries for onsets/offsets of user behavior. It is important to consider other methods of segmentation that would take into account HVS events comprised of the basic eye movements and their properties. Additional research should be conducted that incorporates other information about eye movements and their scanpaths, possibly involving the convex hull area formed by fixation points, inflections in the scanpath formed by fixation points, and the length/duration of the scanpath.

7 Conclusion

Given recordings of users trying to accomplish a specified task via a graphical user interface, this work presents an automated eye movement-driven approach that allows identifying time intervals in which users experience difficulties in locating the interface components necessary for task completion. The identification of such time intervals represented by excessive visual search, aims at reducing the amount of time necessary for the manual inspection of usability issues by examining only "troubled" regions of the recordings.

Several algorithms for the identification of excessive visual search were proposed and the classification performance of these algorithms was compared to a baseline received by manual classification of the recorded data. The results indicate that the accuracy of automated classification varies between 51-61%, while erroneous classification of non-excessive search intervals as excessive does not exceed 26-46%. Further work is necessary to increase classification accuracy and reduce misclassification errors. Additional work is required to investigate the performance of the proposed algorithms on other types of graphical user interfaces and task assignments.

Acknowledgement. This research was funded in part by Emerson Process Management (www.EmersonProcess.com), an Emerson business.

References

1. Andre, W.K., Elizabeth, M.B.: Low-Cost Rapid Usability Engineering: Designing and Customizing Usable Healthcare Information Systems. ElectronicHealthcare 5(2), 98–102 (2006)
2. Andreasen, M.S., Nielsen, H.V., Schroder, S.O., Stage, J.: What happened to remote usability testing?: an empirical study of three methods. ACM, New York (2007)
3. Duchowski, A.: Eye Tracking Methodology: Theory and Practice. Springer, Heidelberg (2007)

4. Goldberg, J., H., Stimson, M.J., Lewenstein, M., Scott, N., Wichansky, A.M.: Eye tracking in web search tasks: Design implications. In: Proceedings of the Eye Tracking Research and Application Symposium, pp. 51–58 (2002)
5. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. International Journal of Industrial Ergonomics 24(6), 631–645 (1999)
6. Ivory, M.Y., Hearst, M.A.: The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. 33(4), 470–516 (2001)
7. Just, M.A., Carpenter, P.A.: Eye Fixation and Cognitive Processes. Cognitive Psychology 8, 441–480 (1976)
8. Komogortsev, O.V., Gobert, D.V., Jayarathna, S., Koh, D., Gowda, S.: Standardization of Automated Analyses of Oculomotor Fixation and Saccadic Behaviors. IEEE Transactions on Biomedical Engineering 57(11), 2635–2645 (2010)
9. Marshall, S.P.: The Index of Cognitive Activity: measuring cognitive workload. In: Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, 2002, pp. 75–79 (2002)
10. Mueller, C., Tamir, D., Komogortsev, O., Feldman, L.: An Economical Approach to Usability Testing. In: IEEE International Computer Software and Applications Conference (COMPSAC), pp. 124–129 (2009)
11. Thyvalikakath, T.P., Schleyer, T.K.L., Monaco, V.: Heuristic evaluation of clinical functions in four practice management systems: A pilot study. J. Am. Dent. Assoc. 138(2), 209–218 (2007)
12. Tobii technology, `http://www.tobii.com`

A Quantitative Evaluation on the Software Use Experience with Electroencephalogram

Hitoshi Masaki[1], Masao Ohira[1], Hidetake Uwano[2], and Ken-ichi Matsumoto[1]

[1] Graduate School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara, Japan
[2] Department of Information Engineering, Nara National College of Technology
22, Yata, Yamatokoriyama, Nara, Japan
{hitoshi-m,masao,matumoto}@is.naist.jp,
uwano@info.nara-k.ac.jp

Abstract. In usability testing, experimenters need to perform a pre-training, so as to control software-use experiences of subjects. The pre-training in usability testing is very important because subjects' software-use experiences have a large effect on a result of a subjective evaluation of software. This paper aims to evaluate the software-use experiences quantitatively using EEG. We have conducted experiments to observe the relationships between subjects' software-use experiences and EEG in using software. As a result, we found that there was a significant difference between them.

Keywords: EEG, Use Experience, Quantitative Evaluation, Usability Testing.

1 Introduction

In recent years, software systems have become much more functional but complicated. Users need usable interfaces to use software systems effectively. In order to develop such the interfaces, evaluating usability has been very important. Existing methods for usability evaluation include interviews, think-aloud protocols [1], questionnaires [2], and so on. These usability evaluation methods extract usability problems from subjective evaluations by users (subjects). They are widely used because they require no special measurement apparatuses and allow usability experts to know usability of software in a relatively easy way if software is completed.

However, we have to carefully consider that the differences of software-use experiences among users have a large effect on a result of subjective evaluation by users. In general, the mental workload of users in using a software system becomes large if the users do not sufficiently learn how to use the software. As a result, users sometimes have negative impressions on the software. Therefore, before usability evaluation, users are trained to learn usage of software in order to make the differences of software-use experiences among users constant. Even so, it is still difficult to adjust the mental workload and/or software-use experiences among users at a certain level since a capacity for learning software is large different among the users.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 469–477, 2011.

Toward constructing a method to reduce the effect of software-user experiences on a result of usability evaluation, in this paper we would like to see whether we can measure software-use experiences quantitatively. As an approach to quantitative measurement of the software-use experiences, we use electroencephalogram (EEG) after users used software. In general, components of the alpha rhythm and the beta rhythm in EEG change by mental activities such as mental calculations, state of tension, condition of excitation, and so on. EEG measurement has some advantages as compared with other bioinstrumentations for the human central nervous system: EEG can be measured by using relatively cheap instruments [3], it imposes fewer limitations in measurement, it does not disturb subjects in using computers, and so forth.

2 Related Work

Measurable information to observe biological activities in the central nervous system includes EEG, magnetoencephalogram (MEG), functional magnetic resonance imaging (fMRI) and so on. In this paper, we use EEG as an indicator to measure the software-user experiences quantitatively. EEG is sometimes used to measure a psychological state of human. Xiaowei et al. [4] set up a learning system on a web site, and recorded EEG of learners visiting the web site. They found that the power of alpha rhythm of users who is concentrating in tasks is lower than that of user who is not distracted. Schier [5] recorded brain waves during a driving simulation task, using a driving simulator. The results showed that the power of alpha rhythm increased when the attention level of drivers decreased. Matsunaga et al. [6] developed a brain wave measurement system for evaluating one's satisfaction and validated the hypothesis that people feel comfortable if the amount of information processing in the brain is small, while people feel uncomfortable if the amount of information processing in the brain is large.

We also measure the power of alpha rhythm and beta rhythm for observing the physiological state of subjects. So our experimental results are easy to compare with the implications and insights from previous work.

3 Experiment

3.1 Overview

To investigate the relationship between software-use experiences and EEG, we have conducted two experiments where subjects used Microsoft Excel 2007. In one experiment, we recruited a group of ten subjects in 2007. Most of the subjects were inexperienced in using Excel 2007 because Excel 2007 had been a brand-new product. Most of them were still familiar with older version, Excel 2003. In another experiment in 2010, we also recruited a group of ten subjects who often used Excel 2007. All the subjects in our experiments were recruited from the graduate school of information science, Nara institute of science and technology. Table 1 shows subjects' usage frequency of Excel 2007. Each subject performed four kinds of tasks (eight tasks in total) described later. We measured EEG of each subject after s/he completed each task. We also asked the subjects to fill in a questionnaire on usability of Excel 2007.

Table 1. Subjects' usage frequency of Excel 2007

Usage Frequency	Experiment in 2007	Experiment in 2010
never	6	0
several times per year	1	2
several times per month	2	3
several times per week	1	5

3.2 Task

Table 2 shows a list of tasks used in the experiments. All the tasks can be performed on both Excel 2003 and Excel 2007, but we expected the group in 2007 confused with the tasks because the user interfaces were very different between Excel 2003 and Excel 2007. Under this experimental setting, the group of the subjects in 2010 should obviously have more positive impressions on Excel 2007 than that of the group in 2007 if software-use experiences have an effect on impressions of software usability.

We gave each subject a data file to make a grade report, which was used in all the tasks. Each subject could spend five minutes to complete a task. We counterbalanced the order of the tasks to minimize learning and remembering effects. The following are the details of each task.

Same Place Task. A subject selects a command with a same name and same position between the two versions of Excel. This task is completed when the subject selects the command.

Different Place Task. A subject selects a command with a same name and different position between the two versions of Excel. This task is completed when the subject selects the command.

Table 2. Task list used in experiment

Task Type	Task Name	Description
Same Place	Open Clip Art Pain	Open clip art pane to select clip art from a list.
	Filter Setting	Set options for data filtering.
Different Place	Display of Version Information	Display the version information of Excel.
	Record of Macro	Run a recorder of macro.
Same Interface	Format Cells	Change date formats from Mar-01 to 03/01.
	Page Orientation	Change a page orientation to landscape and set margins.
Different Interface	Conditional Formatting	Indicate cells that have less than 60 or "Absence" as red font.
	Insert Bar Chart	Insert stacked bar chart of student's scores with chart/axis titles.

Same Interface Task. A subject uses a dialog box with a same composition between the two versions of Excel. In this task, menu name and position were given to the subject before the task. This task is completed when the subject achieved the given operations.

Different Interface Task. A subject uses a dialog box with a different composition between the two versions of Excel. As well As the Same Interface Task, menu name and position were given to the subject before the task. This task is completed when the subject achieved the given operations.

3.3 Environment

Emotional Spectrum Analysis System ESA-16 was employed to record EEG of subjects. After the task, we recorded subjects' EEGs for two minutes at 200Hz sampling frequency in eye-closing, resting condition. Electrode locations are based on the International 10-20 System, shown in Figure 1. We adapted Referential derivation to observe the EEG, and used right earlobe (A2) as reference electrode. As ground electrode, center of the forehead (Fpz) was employed and center of the parietal (Pz) was used as exploring electrode to minimize electromyogram (EMG) artifact. We also recorded electrocardiogram (ECG) from both arms. To reduce artifacts in the measurement, we used headrest and elastic net bandage to fix electrode placed on the head. Before the first task, each subject adjusts height of chair and position of mouse/keyboard.

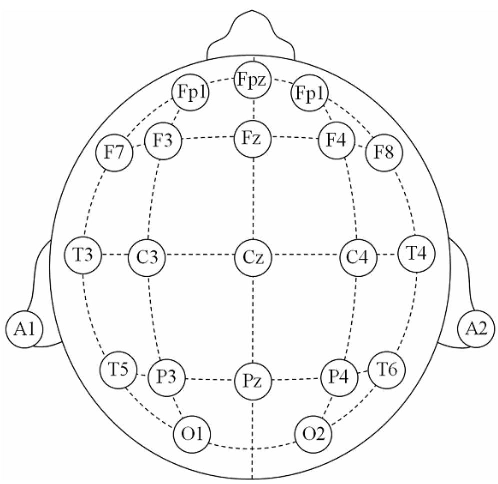


Fig. 1. Electrode locations in International 10-20 System

3.4 Questionnaire

After the all tasks, subjects were answered a questionnaire sheet for investigating subjective satisfactions of Excel 2007 and the usage frequency of each function that was used in the tasks. The questionnaire was created by the authors based on the

Questionnaire for User Interaction Satisfaction (QUIS). Each question about the usage frequency consists of four-point scale (from “Never” to “Few times per week”) and seven-point scale (from “Strongly disagree” to “Strongly agree”) items for evaluating subjective satisfactions.

4 Analysis for EEG Data

We applied power spectral analysis to EEG data we collected at a sampling frequency of 200Hz. Firstly we cut out each subject’s EEG data for 81.96 seconds in eye-closing and resting condition after each task. Next, the target data was filtered to reduce the artifacts from eye blinking, myoelectric activity and so on. We used a high-pass filter (HPF, 3Hz cutoff frequency, +6*dB/oct* attenuation factor), a low-pass filter (LPF, 60Hz cutoff frequency, -6*dB/oct* attenuation factor), and a band-elimination filter (BEF, 60Hz central frequency, 47.5Hz~72.5Hz stopband, second order). The band-elimination filter was used to remove the influence of an alternating-current power supply. After the EEG data was multiplied by the Hamming window and processed with the fast Fourier transform (FFT), we obtained the power spectrum from the EEG data. From the obtained power spectrum, we calculated the respective proportions of alpha rhythm and beta rhythm to all brain waves, and also calculated beta/alpha, which divided the proportion of alpha rhythm into the proportion of beta rhythm. In accordance with classification of the international 10-20 system, we set frequency components of alpha rhythm and beta rhythm to 8~13Hz and 13~30Hz respectively. We also set the range of all brain waves to 3~30Hz.

The proportions of alpha rhythm and beta rhythm to the entire brain waves are often used for observing various activities in the brain [7, 8]. We also use the beta/alpha as indicators for measuring the physiological state of subjects after the tasks. Since the proportions and intensity of alpha rhythm and beta rhythm vary from individual to individual, comparisons of brain waves with the absolute value would be inappropriate. In this paper, we normalized the EEG data from each subject by an average value of each subject’s power spectrum and compared it with the each EEG data.

5 Results

5.1 Correlation between EEG and Questionnaire Result

Table 3 shows correlations between each indicator of EEG and each questionnaire item. This table shows that there were significant correlations between each indicator of EEG and each questionnaire item.

5.2 Relationship between EEG and Usage Frequency

Figure 2 shows the median of beta/alpha by each usage frequency. The vertical axis is the normalized power spectrum of the EEG data. If the value is higher than 1.0, it means the mental workload of the subjects is large. In contrast, if the value is lower than 1.0, it means that the mental workload is small. As shown in Figure 2, the median of beta/alpha of the subjects who often use Excel 2007 (several times per year/month/week) was lower than 1.0. The median of beta/alpha of the subjects who never used Excel 2007 was higher than 1.0. The correlation coefficient between the values of beta/alpha and the usage frequency was -0.599 and there was a significant difference between them (pearson's r, p<0.01).

Table 3. Correlation between EEG and questionnaire result

		Simple to Use	Productivity	Clarity of Information	Pleasant Interface	Easy to Use	Degree of Satisfaction
beta/alpha	pearson's r	-0.616	-0.654	-0.502	-0.672	-0.613	-0.645
	p-value	0.004	0.002	0.024	0.001	0.004	0.002

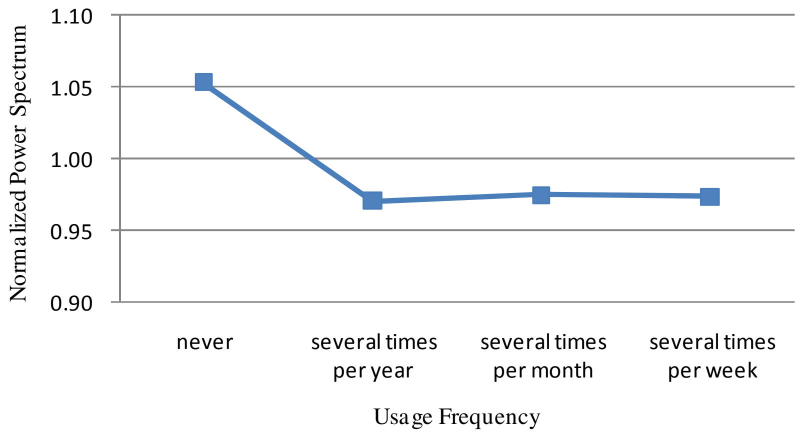


Fig. 2. Normalized value of beta/alpha and usage frequency

5.3 Differences of EEG by Tasks

Same Place/Interface Task. In the experiment in 2007, we hypothesized that the median of beta/alpha would be lower than 1.0 because knowledge and experiences of using Excel 2003 could be applicable to perform Same Place Task and Same Interface Task of Excel 2007 which used menus and dialog boxes as same as Excel 2003. As a result, beta/alpha in Same Place Task was lowers than 1.0. However, beta/alpha in Same Interface Task was over 1.0, contrary to our expectation.

In the experiment in 2010, we hypothesized again that the median of beta/alpha in both Same Interface Task and Same Place Task would be lower than 1.0 because the subjects in the experiment in 2010 were well accustomed to using Excel 2007. Comparing the results in 2007 and the results in 2010, we would like to show the reason why our hypothesis was not supported in 2007.

Figure 3 and Figure 4 show comparisons of beta/alpha in the two experiments. In Same Place Task, it is not difference between the median of beta/alpha in the two experiments. In Same Interface Task, beta/alpha in the second experiment was lower than the beta/alpha of the first experiment, and there was significant difference in beta/alpha between the two experiments (t-test, p=0.012). This indicates software-use experiences have an effect on beta/alpha.

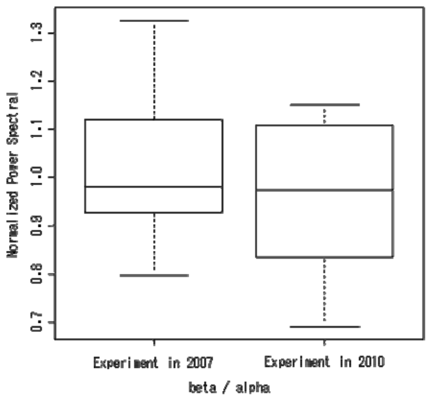


Fig. 3. Difference of beta/alpha at Same Place Tasks

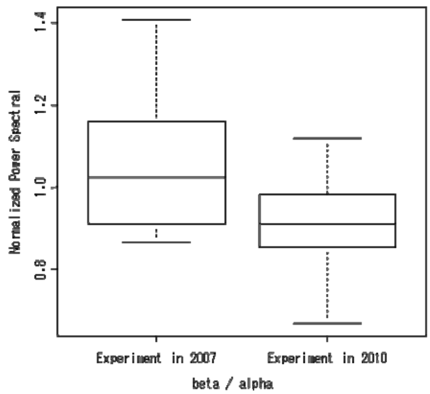


Fig. 4. Difference of beta/alpha at Same Interface Tasks

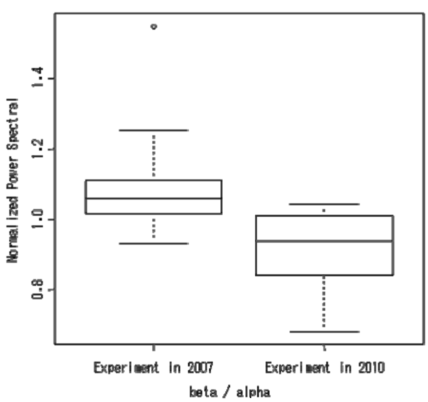


Fig. 5. Difference of beta/alpha at Different Place Tasks

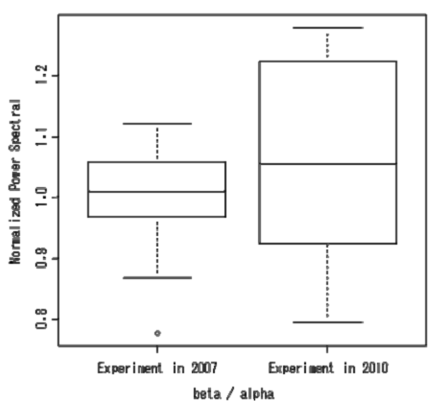


Fig. 6. Difference of beta/alpha at Different Interface Tasks

Different Place/Interface Task. In the experiment in 2007, we hypothesized that the median of beta/alpha would be higher than 1.0 because knowledge and experiences of using Excel 2003 could not be applicable to perform Different Place Task and Different Interface Task of Excel 2007 which used different menus and dialog boxes from Excel 2003. As a result, beta/alpha in both Different Place Task and Different Interface Task was higher than 1.0 as we expected.

In the experiment in 2010, we hypothesized that the median of beta/alpha would be lower than 1.0 because of the same reason in Same Place/Interface Task.

Figure 5 and Figure 6 show comparisons of beta/alpha in the two experiments. In Different Place Task, beta/alpha of the experiment in 2010 was lower than that of the experiment in 2007, and there was a significant difference in beta/alpha between the two experiments (t-test, p=0.041). In Different Interface Task, the beta/alpha of the two experiment were higher than 1.0.

6 Discussion

From the results of the two experiments, we found there is a significant correlation between EEG and subjective evaluation of Excel 2007. The experienced subjects seemed to complete the tasks easily and rated Excel 2007 higher than the inexperienced subjects. Thus, the software-use experiences have a large effect on the subjective evaluation. We also found that there was a significant correlation between beta/alpha and use experience of Excel 2007. The median of beta/alpha of the experienced subjects was under 1.0. In other hand, that of the inexperienced subject was over 1.0. From these results, we conclude that the software-use experiences can be measured by analyzing beta/alpha of subjects.

We would like to discuss the reasons that beta/alpha of Same Interface Task and the Different Interface Task was different from our hypothesis. In Same Interface Task, there was a significant difference in beta/alpha between the two experiments. However, there were not significant differences in the each questionnaire item between the two experiments. Although the usage frequency of Excel 2007 was high, the subjective evaluation of Excel 2007 did not change. We thought that the time to open a dialog box has an effect on EEG. In Different Interface Task, beta/alpha of the experiment in 2010 was higher than 1.0 as well as the experiment in 2007, and there was no significant difference in beta/alpha between the two experiments. However, there were significant differences in the questionnaire items of “Productivity” and “Pleasant Interface” between the two experiments. Furthermore, the all questionnaire results of the second experiments were higher than the all questionnaire results of the first experiments, but there were not significant differences in the questionnaire results of “Simple to Use”, “Clarity of Information”, “Easy to Use” and “Degree of Satisfaction” between the two experiments. From these questionnaire results, we consider that the subjects in 2010 have been accustomed to using the interface in Different Interface Task. We also consider that the interface in Different Dialog Task had a problem on interactivity and it resulted in over 1.0 beta/alpha value.

7 Conclusion and Future Work

In this paper, we had conducted the two experiments to investigate whether we can measure software-use experiences quantitatively or not. As an approach to quantitative measurement of the software-use experiences, we used electroencephalogram (EEG) and analyzed the relationships between EEG and software-use experiences. As a result of the analysis, we found that beta/alpha of experienced subjects become lower than 1.0 while that of inexperienced subjects exceed 1.0. From the findings, we conclude that beta/alpha can be an indicator for measuring the software-use experiences quantitatively.

In our future work, we would like to clarify how long do users spend to accostom use of software, comparing short time use at regular intervals with continuous, long time use. We also would like to show EEG measurement is effective to evaluate *learnability* of software.

References

1. Ericsson, K.A., Simon, H.A.: Protocol analysis: Verbal reports as data. MIT Press, Cambridge (1993)
2. Chin, J.P., Norman, K.L., Shneiderman, B.: Subjective user evaluation of CF PASCAL programming tools. Technical Report (1987)
3. Lee, J.C., Tan, D.S.: Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. In: Proceedings of the 19th ACM Symposium on User Interface Software and Technology, pp. 81–90 (2006)
4. Xiaowei, L., Bin, H., Tingshao, Z., Jingzhi, Y., Fang, Z.: Towards Affective Learning with an EEG Feedback Approach. In: Proceedings of the First ACM International Workshop on Multimedia Technologies for Distance Learning, pp. 33–38 (2008)
5. Schier, M.A.: Changes in EEG alpha power during simulated driving: a demonstration. International Journal of Psychophysiology 37, 155–162 (2000)
6. Matsunaga, H., Nakazawa, H.: A Study on Human-Oriented Manufacturing System (HOMS) – Development of Satisfaction Measurement System (SMS) and Evaluation of Element Technologies of HOMS using SMS. In: Int. Conference on Manufacturing Milestones Toward the 21st Century, pp. 217–222 (1997)
7. Hjelm, S.I., Browall, C.: Brainball - using brain activity for cool competition. In: Proceedings of NordiCHI 2000 (2000)
8. Goo, J.J., Park, K.S., Lee, M., Park, J., Hahn, M., Ahn, H., Picard, R.W.: Effects of Guided and Unguided Style Learning on User Attention in a Virtual Environment. In: Pan, Z., Aylett, R.S., Diener, H., Jin, X., Göbel, S., Li, L. (eds.) Edutainment 2006. LNCS, vol. 3942, pp. 1208–1222. Springer, Heidelberg (2006)

Classification of Interactive System Components Enables Planning Heuristic Evaluation Easier

Llúcia Masip, Marta Oliva, and Toni Granollers

Department of computer science and industrial engineering
University of Lleida, Spain
{lluciamaar,oliva,tonig}@diei.udl.cat

Abstract. Nowadays, new technology continually is turning up and it has incorporated different interactive systems. In a parallel way, people use these technologies and they have to use IS to resolve their tasks without taking into account technology complexities. We consider that the user experience includes a lot of different paradigm that are able to provide users with a positive experience. In this context, our work is focused on enhancing one of the most used usability evaluation techniques. It is an inspection technique that allows carrying out usability reviews without the need of end users. It is called heuristic evaluation. The most difficult task in the heuristic evaluation method is choosing the most suitable set of heuristics for specific interfaces because usability experts have to consider all heuristics and they have to choose the best heuristics for a specific interface. So, the expert evaluator has to know all heuristic and all parts of interactive systems perfectly to find a closed set of heuristics. To make this step easier , we consider that if we have an interactive system component classification, we will be able to detect what components have our interactive system and, to choose the best usability heuristic for each interactive system component. Therefore, we present IS categorization that we will use to decide what heuristics are the best for specific IS according to the main aims of heuristic evaluation.

Keywords: interactive system components, heuristic evaluation, usability, classification, user experience.

1 Introduction

Nowadays, new technology continually is turning up and it has incorporated different interactive systems (IS). In a parallel way, people use these technologies and they have to use IS to resolve their tasks without taking into account technology complexities. So, IS should be easy to use and IS should create positive feelings and comments to achieve a high satisfaction. Therefore, when users use an interactive system, they should feel a positive user experience.

We consider that the user experience includes many different facets that are able to provide users with a positive experience. Nowadays, different facets have appeared to consider user experience in interactive system design like emotions [1], communicability [2], playability [3], cross-cultural [4] among others. But the oldest and most

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 478–486, 2011.

used facet that user experience experts consider when they have to design an interactive system is usability like effectiveness, efficiency and satisfaction that a specific user achieves when he or she uses a specific interactive system in a concrete context of use [4].

Then, usability is an IS feature related to the easiness or difficulty when people use an interactive interface. And, the Human-Computer Interaction discipline provides, among others, a complete set of techniques to evaluate the usability of IS. These techniques are classified in inspection, inquiry and testing techniques [5] [6].

We carried out an exhaustive search for finding user experience evaluation, but we could not find methods including more than one user experience facet. So, we based it on one of the most used methods to assess the oldest user experience facets. It is called heuristic evaluation. It is an inspection technique that allows carrying out usability reviews without the need of end users [5].

Although we based it on the heuristic evaluation method, we wanted to go a step further. If we want to evaluate user experience in an interactive system, we obviously should include end users in the evaluation process. Additionally, we do not want to forget the first use of heuristics. The first heuristics that appeared were advice, or guidelines, to be followed when designing a specific interactive interface. Thus, only with these two premises we detected three different user profiles where each user profile needs to use heuristics in a different way. These profiles are designers, end user evaluators and expert evaluators.

Thus, we based it on the heuristic evaluation process to assess user experience but including different user profiles that will be able to score the heuristics for a specific interactive system.

In this context, our work is focused on enhancing the first step of one of the most used usability evaluation techniques.

To perform an HE it is necessary to complete three steps: planning the evaluation, scoring the heuristics and extracting results. Even knowing that our main and global objective is semi-automation of the HE process, in this paper we have concentrated on the first step: planning the evaluation.

Planning an evaluation implies (i) choosing evaluators, (ii) fixing the most suitable set of rules for the specific interface and, (iii) determining the severity factors to decide on heuristics marks. The most difficult task in this first step is choosing the most suitable set of heuristic for a specific interface. It is the most difficult task because usability experts have to consider all heuristics and they have to choose the best heuristics for a specific interface. So, expert evaluators have to know all heuristics and all parts of an interactive system perfectly, because they have to be able to find a closed set of heuristics that cover each usability feature of each interactive system part.

To make this step easier, we consider that if we have a classification of interactive system components, we will be able to detect what components have our interactive system and, we will be able to choose the best usability heuristic for each interactive system component.

Therefore, we present IS categorization that we will use to decide what heuristics are the best for specific IS according to the main aims of HE. In section 2, we present two real experiences that justify the need of improving heuristic choices. Then, we show related classifications of interactive system components presented by other researchers and next, we explain our classification of hardware and software components. And in addition we detail a discussion about it. Finally, conclusions and future works in this research area are presented.

2 Real Experiences

In this section we show two real usability evaluations in very different interactive systems that reveal difficulties that we had when we chose the best heuristics for each interactive system.

A year ago, the Lleida City Council wanted to evaluate the usability of two new public interfaces that offered services to their citizens. In this section, we want to present both experiences that we were involved in and, then, our motivation for carrying out this work.

The first experience concerns the usability analysis of a virtual website assistant. Second is about the usability and accessibility analysis of interactive physical devices called PIC (Citizens Information Point). The following sub-sections delve into the details of both evaluations.

2.1 Virtual Assistant

Berta is a new virtual assistant provided by the Lleida City Council in its website (http://www.paeria.es). Its main function is to help users to find information about different online procedures that citizens can do in Lleida using its website.

The Computer Department of the Lleida City Council hired us because they were worried about the usability level of their virtual assistant. After starting, those responsible warned us that they can change assistant faces and dialogs used but they cannot modify the interface codification, because it is an external application.

So, they were not interested in usability problems concerning the interface design. Their main interest was about facial expression problems and dialog problems. Taking into account that facial expressions are connected with the dialog used and all this depends on the vocabulary used for end users.

With this information, we studied their pre-requirement and we were aware that we could not use the same heuristic that we used in a common website. We had a different interactive system with different features and our usability goal was completely different from previous experiences that were a good challenge for us.

Finally, we decided to perform a heuristic evaluation ruling out common heuristics. We did not consider it because it did not cover the main objectives of usability evaluation and we considered that the most popular heuristics were not suitable for corporal expressions, dialog and vocabulary used.

So, we reviewed literature to collect heuristics for this type of interactive system and we did not find a concrete set of heuristics which covers the usability features commented above. However, we found four sources that we could mix up to achieve a complete set of usability heuristics for covering all features for virtual assistants. These works are about question-answer systems [7], dialog management in virtual assistants [8], best practices for speaking avatar design [9] and emotional heuristics [10]. Although these works do not present heuristics themselves, we could extract information to consider it for creating heuristics for our concrete features in the virtual assistant.

2.2 Public Kiosks

The second real experience was the usability and accessibility evaluation of interactive public kiosks called PIC (Citizen Information Point) [11]. PIC is a physical device located in public buildings spread throughout the city.

Every PIC allows achieving a large set of public procedures such as obtaining certificates or reviewing personal and local information. PICs are also provided with a printer, keyboard, mouse, screen and digital signature reader through cryptographic card, electronic identity card or USB flash drive. It is like a cash machine, but with computer capability and redesigned elements for those specific locations and situations.

In this case, the main goal was not interfaces that we could see on screen. The main aim was to evaluate usability and accessibility according to their specific localization and their particular physical features.

Thus, we saw that we could not use the most used heuristics (like Nielsen's) because these heuristics, in the same way like experience above, did not cover our evaluation goals and all the features of PICs.

In this case, we used many works to collect a complete set of usability and accessibility heuristics for covering all features of this kind of interactive system. We considered literature about design and evaluation in public kiosks [12] [13] [14] [15] and design and evaluation in general [16] [17]. We also considered notable information extracted from some Internet blogs with confirmed references such as [18] [19]. Apart from these references, we mainly used guidelines of current regulation about digital accessibility such as UNE139801: 2003 and UNE139802: 2003 [20] [21].

With all these articles we could collect a large set of heuristics to evaluate the usability and accessibility of our public kiosk.

2.3 Real Experiences Discussion

As we have seen, every IS has its own features and usability goals widely differ between common websites and virtual assistant interfaces or physical devices. The first aspect we noticed was that we could not use the most used heuristics (like Nielsen's) in both experiences because these heuristics did not cover our evaluations aims and all features of Berta and PICs. Furthermore, the bibliography studied did not provide us specific rules to guide our studies, and then our first difficulty was how awkward it was to choose the most suitable heuristics for those cases. There is no formal and specific definition about usability heuristics for "any kind of IS", and that was our first goal to achieve.

We started searching as much as possible the usability heuristic definitions existing up to now. It is obvious that a lot of researchers have worked to define and modify heuristics concepts. We reviewed all general usability heuristics defined since 1986 and we detected 16 repeated categories. In addition, we concluded in this previous work that 16 categories are applicable in simple websites or desktop applications. To sum up, we consider that 16 categories have to adapt to cover all possible usability aspects that we can find in each interactive interface [22].

Once we have the usability categories, we need to adapt all of them to every kind of IS. Additionally, HE sometimes has specific goals like purchasing, localization or hardware devices, among others. Therefore, it is clear that usability experts have to begin the choice of heuristics determining what parts of IS they want to consider in HE. Even so, if usability experts detect what parts they want to evaluate, the way to choose the best heuristic will become easier, and we justify this in next section.

3 Interactive Systems Classification

What was the way we followed to choose the best heuristics for virtual assistants and public kiosks? We had a large set of heuristics cited in [22] and we reviewed all of them to decide if we could use them to assess some part of our new interactive system. Apart from this review, one of the times we wasted the most was when deciding what parts or components should be chosen for a usability evaluation of those systems. So, we spent time to make some efforts to detect public kiosk and virtual assistant components. And, obviously, we considered common websites in the components search. We achieved a classification of interactive systems components that permitted us to know what parts of the interactive system should be assessed and chose the best heuristics for each part. To sum up, this classification makes planning heuristic evaluation easier.

As a result, we proposes a classification of IS components. We divided IS into two big groups: software IS and hardware IS. We use this classification because we consider that software and hardware systems have different features or components that involve different types of specific heuristics. So, we presented a classification for software and hardware interactive interfaces. Then, some categories (but not all of them) are divided in more components to allow a complete classification of IS components in order to facilitate their evaluation. In the following sections, we detail more about interactive system classification. It is based on website and avatar parts to detect software classification and public kiosk components to show hardware classification.

3.1 Related Classifications of Interactive Systems Components

We did an exhaustive search in literature to find any classification of interactive system components but, unfortunately, we did not find many works about it.

We have highlighted elements detected by Jakob Nielsen and Kara Pernice in their new book "Eyetracking web usability" [23]. The book presents fundamental website design elements in chapter 5. Their list of elements permits designers to consider all the parts in a website design but they do not specify enough to consider all of the elements in our classification. We consider that some of elements presented in the book can be used in a further step for including these elements in our classification of interactive system components. For example, in the book, the shopping cart is considered fundamental in a website design. So, as a future work, we will consider shopping cart implies forms, numerical values, nomenclature, icons and information validation, among others as components in our classification.

Another work related to the classification of interactive system components is [24]. In this work we found a classification of parts of videogames. The way that the author developed his idea is very interesting for achieving a conceptual model of the parts detected in this type of interactive system. But this model only includes specific interactive systems and we wanted to cover a larger set of them.

So, we did not find a classification that includes specific parts of the most used interactive system such as common websites, virtual assistants or public kiosks. In the next section we present a classification of interactive system components including all parts of these interfaces.

3.2 Software Interactive Systems

We consider that software interactive systems are the systems that include some type of software such as websites, desktop applications, video recorders, and mobile phone interfaces. But in addition, we should consider all hardware systems that also include software parts (although this software part will be a little part of a big hardware system). For instance, public kiosk or some electrical appliance, among others, have important hardware components but they also provide us with a little software interface that we cannot forget when we want to carry out a complete user experience evaluation.

According to our past evaluations of common websites, virtual assistants and desktop applications, we can detect some components that are susceptible to a user experience evaluation. So, our classification of software interactive system components shows a complete set of software components that we can detect in many types of interactive systems. To sum up, we present in the next table a classification of software interactive system components that we can use to detect each part of most software interactive systems.

Table 1. Classification of software interactive system components

General group	Specific feature
Type of content	Forms, tables, lists, dates, times, numerical values, money signs
Information	Pictures, news, graphics, format, text, URL, abbreviations, audio, nomenclature, colours, icons
Data management	Information transmission, sign in form, log in form, updating information, information validation, information recovery
Search	Search form, results of search
Navigation area	Pages, titles, cursor, shortcuts
Emergency exits	

3.3 Hardware Interactive Systems

We know hardware interfaces as some physical devices that you can interact with it. Some examples are elevator panels, public kiosks, video recorders, ovens...

As above, we were faced with the challenge to choose which “hard” components of the public kiosk should be analysed. We highlight that a public kiosk, as a physical system, is a complete interactive system with a large set of components and a software part.

So, hardware interfaces are classified into operative systems, navigator, in/out devices, assistive aids, audio, indicators and help and documentation. Then, every category is divided in more components to allow a complete hardware IS classification in order to facilitate its evaluation.

The hardware interactive system categorization is:

Table 2. Classification of hardware interactive system components

General group	Specific feature
Operating system	
Navigator	
In/out devices	Printer, digital certificate, electronic card, mouse, screen, keyboard
Assistive aids	Icons, assistive hardware
Audio	Volume
Indicators	
Help and documentation	

3.4 Interactive System Components Discussion

Note that in hardware classification we include software components such as operating systems or navigators. It is because hardware systems might provide us with software applications and we consider that the same type of hardware interactive system should include the same version of operating system or navigator to avoid confusion for users when they use these systems.

In addition, we should consider that sometimes, IS includes software components. Therefore, this classification is not a mutually exclusive classification but it is a complementary classification because the hardware and software components of systems might be found in a concrete interactive system such as public kiosks.

It is also obvious that sometimes some heuristics can take part in more than one part of an interactive system. For example, there are some heuristics like "the same things have to be aesthetically equals", so it is applicable in indicators, keys, icons, among others. So, some heuristics are the best for more than one of the interactive system components. But we consider that if we know perfectly which components of our interactive system we want to evaluate or which components should be considered, we will be able to choose the best heuristic for each component more easily.

4 Conclusions

In conclusion, we want to concentrate all importance in IS when we want to do an HE. We have an interface and it is the centre of our minds. Then, we can decide what type of goals we want to consider in evaluation: usability features or components of

IS. And we should be able to choose the best heuristics with both types of evaluation aims.

In this work we presented a classification of interactive system components because our main problem when we carry out a user experience evaluation through a heuristic evaluation in a new interactive system is detecting all interactive system components of our system for choosing the best heuristics for each component. Thus, we consider that if we have a classification and this classification is linked to specific heuristics (this is part of our future work), we will obtain the best heuristics for an interactive system more easily and in a more comfortable way than if we have to review each interactive system, detect all parts, look for the best heuristics and so on.

Finally, we know that this classification is not a closed group. It is the first step in achieving a large and complete classification of all interactive systems.

As a future work, we want to achieve a general and complete classification of interactive system components because we want to be able to detect interactive system components with a general goal of interface. For instance, if we have a website, we will detect that the website includes a main part of privacy and forms because this website is an online shop. So, we think it will be possible if we consider interaction patterns in our project.

Acknowledgments. The work described in this paper has been partially supported by Spanish Ministry of Science and Innovation through the Open Platform for Multichannel Content Distribution Management (OMediaDis) research project (TIN2008-06228) and it has been partially supported by Universitat de Lleida for pre-doctoral fellowship to Llúcia Masip. An acknowledge to Development of Adaptive and Collaborative Systems (DESACO) research project (TIN2008-06596-C02-01/TIN) also must be done.

References

1. Leitner, G., Hitz, M., Melcher, R.: The Role of Usability in the Design and Evaluation of Dynamic Traffic Displays. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 205–220. Springer, Heidelberg (2008), doi:10.1007/978-3-540-89350-9_15
2. Prates, R., de Souza, C., Barbosa, S.: Methods and tools: a method for evaluating the communicability of user interfaces. Interactions 7(1), 31–38 (2000), `http://doi.acm.org/10.1145/328595.328608`
3. González, J.L., Padilla, N., Gutiérrez, F.: From Usability to Playability: Introduction to Player-Centred Video Game Development Process. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 65–74. Springer, Heidelberg (2009)
4. Jiang, O., De Bruijn, O., De, A.: The Perception of Cultural Differences. In Online Self-presentation. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 672–685. Springer, Heidelberg (2009)
5. ISO. International Standard ISO/IEC 9126. Information technology – Software product evaluation – Quality characteristics and guidelines for their use, International Organization for Standardization, International Electrotechnical Commission, Geneva (1991)
6. Nielsen, J., Mack, R.L.: Usability Inspection Methods. John Wiley & Sons, New York (1994) ISBN: 0-471-01877-5

7. Dix, A., et al.: Human-Computer Interaction. Prentice-Hall, Madrid (2009)
8. Martínez, P., Vicedo, J. L., Saquete, E., Tomás, D.: Sistemas de pregunta-respuesta Universidad de Alicante. Departamento de Lenguajes y Sistemas Informáticos (2007) http://rua.ua.es/dspace/handle/10045/2525
9. Rodrigo, L., García, A., Martínez, P.: Planeamiento semántico y pragmático para gestión de diálogos en asistentes virtuales. Procesamiento del Lenguaje Natural, Revista, num. 28 (2002)
10. Sitepal. Speaking avatar best practices guide (2003), allowed in http://www.sitepal.com/pdf/bestpractice_2008.pdf
11. Lera, E., Garreta, M.: Diez heurísticos emocionales. Revista Faz. num 2, pp. 68–71. Revista de Diseño de Interacción. Barcelona, España (2008)
12. Oliva, M., Masip, L., Granollers, T.: Evaluación de usabilidad y accesibilidad de un conjunto de dispositivos interactivos denominados Puntos de Información Ciudadana. Scire: representación y organización del conocimiento 16, 35–46 (2010)
13. Maguire, M.C.: A Review of User-Interface Design Guidelines for Public Information Kiosk Systems. In: Stephanidis, C., Carbonell, N. (eds.) Electronic Proceedings of the 3rd ERCIM Workshop on User Interfaces for All (1997), http://ui4all.ics.forth.gr/UI4ALL-97/maguire.pdf (15/09/2008)
14. Nieto, M.: La usabilidad de los terminales táctiles. Evolucy (2002), http://www.evolucy.com/esp/columns/20020515_usabilidad_tactil.html (15-7-2009)
15. Gutiérrez, M.: ¡Larga vida de los cajeros¡, Revista Faz. November 2007 número 1, http://www.revistafaz.org/numero1/cajeros.pdf (19/09/2008)
16. Evolucy. Evaluación heurística de un quiosco de autoservicio de fotografía digital: caso práctico (2004), http://www.evolucy.com/esp/projects/evaluacion_heuristica_kiosk.html (8-7-2009)
17. Granollers, T., Lorés, J., Cañas, J.J.: Diseño de sistemas interactivos centrados en el usuario. Editorial UOC, Barcelona (2005) ISBN: 84-9788-320-9
18. Sidar, F.: Principios del Diseño Universal o Diseño para Todos (19/09/2008), http://www.sidar.org/recur/desdi/usable/dudt.php
19. Usable y Accesible (19/09/2008), http://olgacarreras.blogspot.com/
20. facilusar.com. (19/09/2008), http://facilusar.blogspot.com/
21. UNE139801. Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Hardware. AENOR. Depósito legal: M 42219:2003 (2003)
22. UNE139802. Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Software. AENOR. Depósito legal: M 42220:2003 (2003)
23. Masip, L., Oliva, M., Granollers, T.: Hacia la semiautomatización de la evaluación heurística: Primer paso, categorización de heurísticas, Interacción 2010, Grupo Editorial Garceta, Valencia, September 7-10, pp. 359–368 (2010)
24. Nielsen, J., Pernice, K.: Eyetracking web usability. New Riders, Berkeley (2010) ISBN 9780321498366 (pbk.) 0321498364 (pbk.)
25. González, J.L.: Jugabilidad: caracterización de la experiencia del jugador en videojuegos. Doctoral Thesis. Universidad de Granada (2010)

Clustering Analysis to Evaluate Usability of Work-Flow Systems and to Monitor Proficiency of Workers

Toru Nakata

Digital Human Research Center, National Institute of Advanced Science and Technology, Aomi 2-3-26, Koto-ku, Tokyo, 135-0064, Japan
toru-nakata@aist.go.jp

Abstract. In order to evaluate usability, we often interview the users with using questionnaire sheets. However, conventional processing methods for questionnaire data are so simple that we cannot mine maximum information from the users' opinions. This paper proposes a new method for deeper analysis of the user opinions. Using the vector quantization method, we can classify users into groups reflecting their skill grade. Also, by observing learning curves of the tasks, we can evaluate hardness of mastering each task and detect the defects of the work-flow to be improved. This paper explains the idea and mechanics of the method with referring an actual example, which is a questionnaire investigation for workers in a real office to examine usability problems on their works. The result of data analysis pointed out several very hard defects in the work-flow system, on which not only the novices but the experts are also facing difficulties. Those very hard defects cannot be solved by experience or training, because the experts cannot cope with them. Thanks to the vector quantization analysis, we can distinguish between difficult points that can be solved by workers' experience and such very hard defects that require drastic reforms for improvement from.

Keywords: Usability testing, vector quantization, questionnaire sheet method, work-flow reform.

1 Introduction

Quality of work-flow design is decisive for the efficiency of the company performance. Badly designed work-flow induces human errors, unnecessary additional costs, and accidents. The reduction of complexity of work-flow is required for various works in our life [1].

There are already analysis methods of work-flow design from the viewpoint of software engineering [2]. In such analysis, the work-flows are described as canonical form like petri net, and then their efficiencies are evaluated almost mathematically.

Clerical works, however, are often unsuitable to such formalistic analysis. Clerical works in actual offices are mainly processed by human workers. Therefore, the efficiency and comfort of the work process largely depend on usability of the tools or the system that the workers use. Usability is a psychological factor [3], [4], [5], so the

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 487–496, 2011.

formalistic analysis of mathematical "work-flow analysis" is not adequate for it. Investigation focused on psychological factors is required.

One of the most difficult things to investigate psychological factors in clerical works is to monitor actual conditions of the works. Investigators of work-flow design are often afforded only a limited number of subjects for usability test, and it is difficult to survey the details of the real condition in many offices in the whole company.

In order to collect the raw and direct opinions from large number of people, we usually use the questionnaire sheet method [6], [7], [8], [9]. But it is doubtful that we have been successful in acquiring plenty information from the questionnaire data with traditional methods. The theme of this paper is how to analyze the data of the questionnaires efficiently and deeply.

The most straight-forward (and therefore most popular) method to handle the data from questionnaire sheets is calculating averages of each inquiry. The analysts warn defects of the design, which got bad reputations. When we want to get deeper information, we calculate correlations among the data and find out some hidden relations. But these analyses are elemental, and most of information contained in the data will be ignored.

Questionnaire data are generally big. The data size equals to the number of the subjects times the number of the questions. That means the data matrix contains very rich information, and the statistical analysis should be more complex to mine such big data.

The purpose of the paper is to produce an analysis method that consists of cross analysis to mine deeper information such as:

- Comprehension of users' proficiency process,
- Categorization of users (figuring user models of novices and experts),
- Diagnose on defects by monitoring learning curves.

The paper will show the way for those purpose by referring an example of applying the vector quantization method to analyze questionnaire sheet data that is a set of workers opinions about their work-flow management.

2 Example of Analysis: Workers and Work-Flow in a Clerical Office

Since the capacity of the paper is limited, we explain the method with referring an application example in parallel.

2.1 The Worker Subjects and Their Work

A survey was carried out in a clerical office of a company. The office mainly deals financial works on papers and computers with facing many visiting customers.

The investigation using questionnaire sheet is planned to evaluate the conditions of the works and to detect real causes of difficulties and inefficiencies.

The questionnaire survey was carried out for 44 workers.

2.2 Designing Questionnaires on Usability and Difficulty of the Work-Flow

Concepts on Design and Data Processing. In general, office workers know to the last detail of the good points and defects of tools, systems, manuals, and flows of their work. So, direct questionnaire investigation for the workers is the best way to find out ways to improve the work.

The questionnaire sheets should have peripheral inquires about background and circumstance besides the main questions on opinions about usability. As shown later, peripheral data about such as personnel experiences and general opinions on working environments often have unexpected correlations toward the opinion data on usability. And then we may find hidden reasons on generation of the good points and the defects of the work-flow.

On designing questionnaire analysis method, we have to be careful about processing questionnaire data with highly complex mathematic techniques. Data of questionnaires tend to be less quantitative. The answers about opinions of the subjects are usually qualitative. To make them more quantitative, we can grade them, for instance, "good," "neutral," and "bad," and deal them as numerical data. Knowing it is not accurate, we transfer the data from the ordinal style into numerical style as "1," "2," "3" and so on. It is just for a practical reason that numerical data are convenient for statistical analyses.

Since this enumeration from qualitative opinion data to pseudo-quantitative data is not linear transformation in general, analytical mathematic methods such as Principal Component Analysis (PCA) and Factor Analysis (FA) are not suitable. Vector quantization method is not strongly analytical, since it switches the approximation functions for each cluster unlike PCA and FA. So it may handle pseudo-quantitative data better.

The Questionnaire Sheet. We provide 49 questions asking conditions and opinions about the difficulties of certain tasks (Table 1).

The answers are replied by filling check-boxes. Each subject chooses one answer choice for each question. Each answer choice is assigned to values that are shown as numbers in the brackets as the following. By regarding these numbers as representative values of the answer choices, the answers are dealt as pseudo-quantitative numerical data.

The answer data of each subject forms a 49-dimensional vector (Fig. 1). We name the first 4 elements (about the subject's experience) as vector L, 5th to 33rd elements as vector D (about subjective difficulties), and 33rd to 49th elements are vector T (about evaluation on the assistance tools).

We employ vector D as the key for clustering the subjects (Fig. 2). We use vector quantization method, which is one of common techniques for clustering, and then classify all the subjects into 4 groups. Each cluster are a group of the workers who gave similar opinions. We grade the groups in respect to proficiency that are evaluated from values of D (vector of difficulty).

Then we will explain the reason of generation of the proficiency differences by observing relationships among data D, L, and T. Finally, we examine the learning curves to rate the hardness of mastering each task.

Table 1. The questionnaire sheet

About experience of the tasks:
Q1. Number of experienced years on the tasks: less than half year (0) / half to one year (1) / one to two years (1.5) / two or more years (3).
Q2. Number of days processing the tasks in a week: rare (0) / less than one day (1) / 1-4 days (2.5) / everyday (5).
Q3. Number of experienced year on reviewing colleagues' works: (answer choices are same to Q1).
Q4. Reviewing frequency: (choices are same to Q2).

Recognition of overall condition of the task executing:
Q5. Number of days in a week when you feel difficulty on processing the works: (choices are same to Q2).
Q6. Frequency of feeling difficulty on reviewing colleagues' works: (choices are same to Q2).

Task 1: Communication with customers and colleagues:
Q7. Filling up the form to represent customer's intention precisely.
Q8. Handing and receiving stuffs to/from customer.
Q9. Explaining circumstance of tasks to colleagues
Q10. Handling and receiving stuffs to/from colleagues
From Q7 to Q10, the answer choices are "(1) easy / (2) somewhat easy / (3) neutral / (4) somewhat difficult / (5) difficult."

Task 2: Filling numerical figures on the form:
Q11. Do you know the most typical error pattern of the task? (1) Yes, (2) No.
Q12. Can you do without the manual? Confident (1) / somewhat confident (2) / neutral (3) / somewhat unconfident (4) / unconfident (5).
Q13. Do you miss filling up necessary data on the forms? Almost never (1) / rarely (2) / sometimes (3) / often (4).
Q14. Can you detect colleagues' errors by reviewing? (Choices are same to Q13).

Task 3: Filling data after processing required arithmetical transformation (i.e. exchanging currency, transferring numbers of elements to number of sets, etc.)
Q15-19 are same to Q11-14.

Task 4: Sealing revenue stamps. (In Japanese law, revenue stamps are basically required for receipts assigned over 30,000 JPY. Besides it, some exceptional cases exist.)
Q20. Do you know the most typical error pattern of the task? (1) Yes / (2) No.
Q21. Can you do without looking up the manual? (Choices are same to Q12).
Q22. Do you know exceptional cases? (1) Knowing detail / (2) roughly knowing existence of problem (3) completely not knowing.
Q23. Do you forget sealing necessary stamps? (Choices are same to Q13).
Q24. Do you misjudge the necessity of the stamp? (Choices are same to Q13).
Q25. Can you find out colleagues' errors when reviewing? (Choices are same to Q13).

Task 5: Giving and receiving signatures on forms
Q26. Do you know the most typical error pattern of the task? (1) Yes / (2) No.
Q27. Can you do without the manual? (Choices are same to Q12).
Q28.Do you miss filling up necessary signatures? (Choices are same to Q13).
Q29. Do you add unnecessary signatures mistakenly? (Choices are same to Q13).

Table 1. (*continued*)

Q30. Do you miss filling up necessary signatures in case of complex works? (Choices are same to Q13).
Q31. Do you miss filling up date or other trivial information around signatures? (Choices are same to Q13).
Q32. Do you skip verification of customers' signatures mistakenly? (Choices are same to Q13).
Q33. Can you find out colleagues' errors when reviewing (Choices are same to Q13).
Evaluation on usability of the forms and the documents
Q34. Size of letters: (1) prefer bigger / (2) keep current / (3) prefer smaller.
Q35. Spatial density: (1) prefer sparser / (2) keep current / (3) prefer thicker.
Q36. Do the forms have unnecessary boxes? (1)No / (2) some / (3) many.
Q37. Do the forms lack necessary boxes? (1)No / (2) some / (3) many.
Q38. Color of letters: (1) good / (2) neutral / (3) bad.
Q39. Do you lose pieces of the forms? (Choices are same to Q13).
Evaluation on usability of the manuals
Q40. Frequency of looking up manuals: (1) often/ (2) sometimes/ (3) rarely/ (4) never used (5) do not know the location of them.
Q41. Easy to look up? (1) Yes / (2) neutral / (3) no / (4) never used.
Q42. Easy to understand terminology? (Choices are same to Q41).
Q43. Easy to find out the procedures that you have to do? (Choices are same to Q41).
Q44. Is it helpful in general? (Choices are same to Q41).
Evaluation on usability of the quick references
Q45-49 are as of Q40-44.

2.3 Analysis of Clusters of the Workers

Correlations among Opinions of the Inquires. Tracing changes of answers of each subject (as drawn lines in Fig. 1), we can find out correlations among the questions. In contrast, we cannot tell such detail structures by just observing averages as in traditional simple analyses.

Clustering Result. Applying the clustering, we found 4 groups of the workers. The meaning of the group can be detected from the parameters concerning their experience (Table 1). We can see the detail differences of data on each questions by grouping the answers in respect to the clusters (Fig. 3). Some questions got similar opinions from all of the worker clusters, and other questions did not. Those questions with diversity of answers means they are concerned with experience of the work.

The dots mean answers of the subjects, the lines connecting the dots represent answers of same subject. There is positive correlation between opinions of inquiry Qi and Qj. Negative correlation exists between opinions of inquiry Qj and Qk. The difference of opinions allows cluster the subjects into groups Cluster 1 and 2. Taking averages of answers for each inquires (shown as broken lines in the middle) is a typical traditional method, but it neither detects such correlations nor clusters.

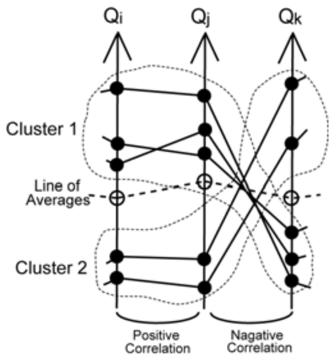


Fig. 1. Tracing each subject over the data to observe correlations

Table 2. Clusters of the workers classified by contents of the opinions

Cluster ID	Relationship to amount of experiment	Number of workers
1 ★	Workers with long experiences on the job.	12
2 ▼	Workers doing the works very frequently.	15
3 ●	Intermediate workers.	10
4 ▲	Novice workers.	7

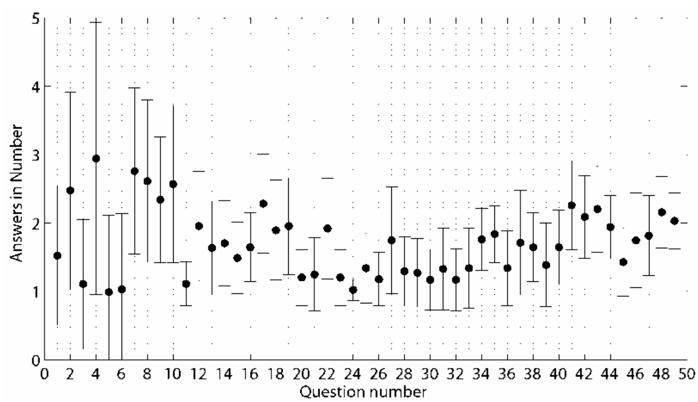


Fig. 2. Means and standard deviations of the answers over all subjects. The answers are transferred to pseudo-quantitative values.

Reasoning the Clusters. When tracing each subject (like shown in Fig. 1), we may find the typical patterns of the answers. By using clustering methods such as vector quantization we are able to nominate the patterns precisely.

In questionnaire sheet investigation about usability, there exist several typical answer patterns with keeping clear separation gaps to other patterns. This tendency appears because the population of the subjects is consisted of different groups, i.e. novices and experts. Difference of proficiency of the workers makes differences on the opinion answers. However, other factors besides the experience length may have influences.

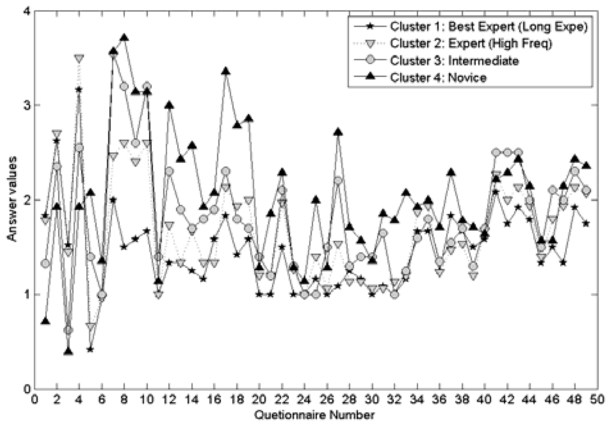


Fig. 3. Average of answers of each cluster

To detect what factors are effective and to check their power, we process the data as 3 steps:

1. Select the parameters which are candidates of cause factors of the cluster separation. (For example, the data about background of each worker should be dealt as a candidate.)
2. Do clustering of the data without using the candidate parameters. Make clusters in respect to only the rest part of the answers, i.e. data about opinions about the workflow.
3. Monitor the correlation between the candidate parameter and difference of the clusters, by tracing answers of each cluster (Fig. 4). If a set of the trace lines from a cluster to a certain candidate parameter converges in unique position, it means that the cluster has strong relationship to the candidate parameter. Narrowness of convergence and separation against other clusters indicate the power of the relationship. Then we can evaluate the candidate parameters.

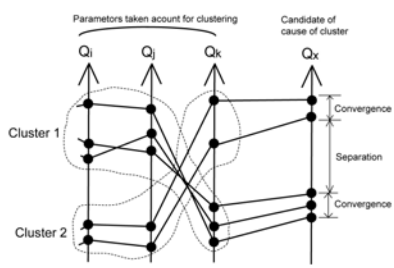


Fig. 4. Detecting the cause of the cluster separation. The cause parameter has strong correlations to difference of the cluster. By tracing the data of each subject (drawn as lines), we can observe the power of influence to the cluster separation.

2.4 Difficulties of the Work and Learning Curves

Concept of Learning Curve Analysis. We can find defects in the work-flow by observing of "learning curves." Learning curves can be drawn by plotting opinion answers about the difficulties of each task in the work-flow in respect to the length of experience which each worker has.

Shape of learning curve reflects condition of education and training for the workers about the task (Fig. 5). Bent learning curves imply mature of proficiency is fast or slow. Slow mature curve indicates that the task is very complex, so that only the experts can do the task. We can find that such difficult tasks must have certain problems on their process and should be reviewed and fixed.

Strange curves may be appears. Normally, the more you have experience, the less you make mistakes. Such monotonic decreasing relationships between experience and hardness of the work are common to normal works. But non-monotonic learning curves sometimes appear. They mean the tasks are suffering disturbance than are caused by certain strange circumstances of the work-flow.

Such strange curves, however, are not rare in actual investigations. There may be several reasons for generation of non-monotonic, and the most plausible reason is that the expert workers are often entrusted with the most difficult tasks. So over-all hardness of the work is stronger than that for the novices and intermediate workers. When the strange curves are found, we should review assistances and special training for over-intermediate workers to process their complex works.

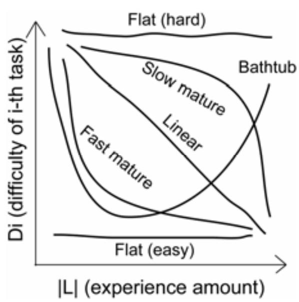


Fig. 5. Some typical patterns of learning curves

Results and Discussion on Learning Curves. We got the learning curves shown in Fig. 6.Non-monotonic curves are found in Q11, 19, 20, 25, 26, 34, 37, 41, and 42. The working conditions concerning thoses questions are thought having special and severe prombles. For instance, non-monotonic learning curves appeared in questions about knowleges on occurrence distrubution of typical clerical mistakes (Q11, Q20, and Q26) and reviewing skill (Q19 and Q25). This means that the experts are facing more difficult tasks and are less confident than the intermediate workers, so that there should be trainings for the experts to improve their skill.

Fig. 6. Learning curves from the novice cluster (▲), the intermediate cluseter (●) to the two expert clusters (★ and ▼). X-axis: experience amount estimated as norm of experience vector L. Y-axis: average of the answers in each cluster. For inquries Q5 to Q49, the larger the value is, the worse the opinions about the inqurie are.

References

1. OECD: Cutting Red Tape — National Strategies for Administrative Simplification. OECD (2006)
2. Wil, v.d.A., Hee, v.K.: Workflow Management. MIT Press, Cambridge (2002)
3. Gilb, T., Weinberg, G.M.: Techniques for Reliable Keyed Input. Winthrop Pub. (1976)
4. Gillespie, R.: Manufacturing knowledge — A history of the Hawthorne experiments. Cambridge Univ. Press, Cambridge (1991)

5. Freiberg, K., Freiberg, J.: Nuts! — Southwest Airlines Crazy Recipe for Business and Personal Success. Bard Press (1996)
6. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. IBM Technical Report 54.786 (1993)
7. Lin, H.X., Choong, Y., Salvendy, G.: A Proposed Index of Usability: A Method for Comparing the Relative Usability of Different Software Systems. Behaviour & Information Technology 16(4/5), 267–278 (1997)
8. Kirakowski, J., Cierlik, B.: Measuring the Usability of Web Sites. Human Factors and Ergonomics Soc. Annual Meeting Proc., Communications (5), 424–428 (1998), Human Factors and Ergonomics Soc.
9. Gersho, G.A., Robert, M.: Vector Quantization and Signal Processing. Kluwer Academic Pub., Dordrecht (1992)

Fundamental Aspects Concerning the Usability Evaluation of Model-Driven Object Oriented Programming Approaches in Machine and Plant Automation

Martin Obermeier, Steven Braun, Kerstin Sommer, and Birgit Vogel-Heuser

Institute of Automation and Information Systems in Mechanical Engineering,
Technische Universität München, Boltzmannstraße 15, 85748 Garching, Germany
{obermeier,braun,sommer,vogel-heuser}@itm.tum.de

Abstract. Within the world of automation the trend of model-driven object oriented (oo) engineering has brought up fundamental questions about the applicability of these programming paradigms for Programmable Logic Controller (PLC) software. The authors present the results of previously conducted experiments on the usability of the classic procedural paradigm (IEC 61131-3) in machine and plant automation compared to model based approaches for PLC programming, in particular Unified Modeling Language (UML) and domain specific modeling languages. Extrapolating these experiments, we propose a way of enhancing usability evaluations by two means: First we present an improved modeling tool. Second, in order to determine the complexity of the tasks required to develop a PLC-program and to create constant boundary conditions for experimental studies, we propose using Hierarchical Task Analysis (HTA) on both model-driven oo and the state of the art programming approach, concerning typical scenarios. Finally the results of our work are discussed.

Keywords: Usability, PLC-Programming, model-driven engineering.

1 Introduction

Our main goal is and was to evaluate the benefit of object orientated paradigms by using UML as modeling notation in automation software, through measuring the increase of efficiency and quality of the programs. This paper answers the following questions: Is a tool supported UML modeling approach useful and beneficial for software engineering in automation in terms of software quality and development efficiency? How can HTA be used for typical industrial scenarios to support the development of realistic, jet significant and controllable experimental designs?

In our research we focus on automation software development for machine and plant automation. The software on Programmable Logic Controllers (PLCs) is typically programmed by using the IEC 61131-3 [1] languages. The IEC 61131-3, defining five procedural programming languages (Structured Text, Sequential Function Chart, Ladder Diagram, Function Block Diagram and Instruction List), is still

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 497–506, 2011.

accepted and applied as a worldwide industry standard in automation software engineering. Over a long period of time we have evaluated the advantages of UML modeling approaches in comparison to the IEC languages. During these evaluations we developed an UML code generator plugin for a standard IEC 61131-3 programming environment (CoDeSys 3.0) [2]. Now we have the possibility to evaluate the model-driven oo programming approaches for PLC systems under more realistic circumstances, by using three UML Diagrams (Class Diagrams, State Charts and Activity Diagrams) with CoDeSys 3.0 [3]. Although the tool is now available and we have basically shown the benefits, regarding the study of the ARC Advisory Group [4], the impact of non-IEC programming approaches on embedded controllers is still minimal. In Order to increase the level of acceptance in industry we are on the way to evaluate more realistic scenarios, which will be presented later on in this paper.

In the following, we will briefly discuss the state of the art in automation software engineering for machine and plant manufacturing industry and the applicability of UML as notation on recent evaluation experiments of UML in PLC programming.

2 State of the Art

Control engineers in automation must cope with challenges in all phases throughout development of control systems. Since existing technologies and paradigms are limited in their effectiveness, a new approach is needed for a higher productivity in modularity and reuse of software components other than on a function block level where it is mostly applied [5], [6].

On this basis, a series of controlled experiments to evaluate the benefit of modeling as such, and more specifically of UML for process modeling and subsequent PLC-programming were conducted.

The UML (current version 2.3) is an established standard for software development. It is specified by the Object Management Group (OMG) and available as ISO/IEC 19501. UML can be used to create models from a software perspective or on a conceptual basis. If the UML is used from a software perspective, object-oriented features can be used for model construction.

2.1 Related Work

Two approaches based on the UML and the Idiomatic Control Language (ICL) were compared to the typical procedure of PLC programming without any notational support [5], [7], [8]. One of the main results stated by the subjects was the fact, that no procedural method for the modeling task was provided. Subjects felt confused by the number of possible diagrams provided by the UML and were insecure about the correct sequence in modeling. Additionally, most of the subjects criticized the lack of tool support for the modeling task.

Based on these results, Katzke specified UML-PA as a sub set of UML [9], [10], [11]. Like the UML, UML-PA is composed out of various sublanguages. These languages describe different structures, interactions and behaviors. By selecting specific language elements from approaches in application software development, UML-PA is a customized modeling language for automation and was developed in DiSPA [12].

The practical applicability of a modeling language is as important as its formal specification. The UML-PA was designed to lower complexity with fewer types of diagrams and modeling elements.

Experiments showed the necessity of a domain specific UML and a restricted number of diagrams, as well as support of a tool and a methodology. Support of legacy code is strongly required. In manufacturing and plant automation many already tested lines of code exist. This legacy software often constitutes valuable know-how and needs to be reused due to economic and quality reasons. Hence programming tools require the ability to present the software structure in an intuitive way. Importing and visualizing existing IEC 61131-3 code in the form of UML class diagrams should be possible. Executable UML models must be mature and complete as a basis for a successful and productive usage and most importantly, accepted by the end users. Our UML generator is embedded into CoDeSys 3.0, the reference implementation of the object-oriented extension of IEC 61131-3. The generator supports three diagram types as there are class diagrams for structural description, state chart and activity diagram for behavioral description. All these diagrams are specified in a complete manner [2] and [13]. The generator provides modeling, coding and online debugging of UML models, where object-oriented elements can be mixed with traditional IEC language elements. Since the debugging functionality is entirely integrated in the IEC 61131-3 environment, the implementation can be monitored during runtime.

More complex design patterns like those already applied in classical application development have not yet been elaborated in automation control systems. It will be an important future research project to derive such patterns in order to reduce development time and cost by raising the factors modularity and reuse.

Industrial projects, their experts, as well as lab experiments should be included to evaluate the benefit in terms of quality and time reduction based on criteria for the usability evaluation of programming notations and programming languages in lab experiments proposed by [14] and [15].

The standard procedure to promote understanding of the procedural knowledge involved in doing a task is conducting a task analysis, which contains a description of the particular action steps carried out when the task is being executed. This can be used for recommending training or for predicting the efficiency when using a certain notation [16]. In our studies, we decided to apply the "Hierarchical Task Analysis" (HTA) [17], which is the original and most common form of task analysis and has been successfully used in many practical applications [18]. In a HTA the task is described in terms of a hierarchy of tasks and subtasks usually represented as a kind of tree diagram. The level of detail may vary from showing only the main tasks till listing every single operation a user has to perform for accomplishing a goal.

Alternatively, the "goals, operators, methods, and selection rules" (GOMS) model may be used to analyze the tasks involved. This method – introduced by Card et al. [19] – is closely related to HTA [18]. It describes a task as a hierarchy of goals and sub goals, methods (consisting of a sequence of basic actions) in order to reach the goal and rules for selecting the appropriate method in a specific situation. The main difference between GOMS and HTA – besides their theoretical background (scientific research on human cognition versus pragmatic common-sense observation – is, that GOMS models represent plans and operations in a uniform format whereas in HTA descriptions the plans appear only as annotations to the sequence of actions [16].

A simplified version of the GOMS model is the Keystroke-Level-Model (KLM) proposed by Card & Moran [20]. This method calculates execution time for an entire task by summing the estimated times of the individual actions, called “operators”. That could be, for example, pressing a key but it also comprises mentally preparing to perform an action and waiting for system response. There have been successful attempts to adopt this procedure to GOMS models as well [19]. However, this does not seem to be a promising approach when comparing modeling notations, because “different logic control design methodologies require substantially different higher level constructs”, as Lucas points out [21]. Lucas [21] tries to avoid that problem by using a higher level model, where the basic actions are not individual keystrokes but functions the user must perform (e.g. “add a module”) [22]. Nevertheless, a major limitation persists: There can merely be obtained a very rough prediction of the real execution time, which is only valid for error-free operating experts but not for real operators who are still learning the system and/or make occasional errors. Furthermore, it does not consider individual differences among users, which can lead to huge performance differences (see previous chapter for details).

A key challenge of every task analysis examining different modeling notations remains their high degree of freedom and, consequently, the huge amount of permissible activity sequences. Until now, there is no best practice concerning this issue.

2.2 Conclusion

Despite its simplicity, Hierarchical Task Analysis (HTA) [18] contains most of the characteristics of even the most complex task modeling approaches. Furthermore, it has the advantage of being widely known and used in the task analysis community. In the following we present several ways of facilitating usability evaluation experiments through HTA support, beginning with the possibility to develop specific tasks for empirical experiments considering the research objectives and previous knowledge of the participants.

3 Possibilities of HTA in Usability Evaluation

In this chapter possible applications of HTA in PLC programming usability evaluations will be proposed.

Extrapolating from previous studies [Ka08], [Fr09] we analyzed the task complexity in the experiment as an aspect for improvement. Depending on the level of task complexity either the error level rises massively with the task complexity and reduces the significance of the results or with a diminishing level of complexity the experiment’s results are no longer comparable to real PLC programming tasks in machine and plant automation. Although extending and intensifying the participants’ training alone could possibly counter part of the effect of rising error rates, the needed time and resources cannot guarantee a gain in significance. Thereby, the balance between realistic tasks (practiced in industrial engineering) and experimental tasks (with regard to operability) should always be considered to get generalizable results. Hence we propose HTA as a possible method to support development of a solvable, jet realistic experimental task as well as the necessary training.

First the generic abstract tasks for solving a PLC programming assignment have to be analyzed. These can be described with tasks concerning the preparation of the implementation (see Fig. 1. "Gather information", "Identify interfaces to the surrounding environment") and the implementation of the system functionality itself.

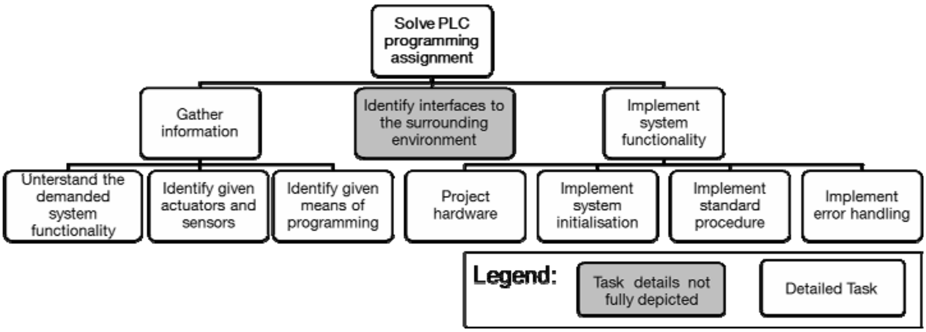


Fig. 1. Abridged graphical version of a Hierarchical Task Analysis for a generic PLC programming assignment

Now a selection of scenarios in PLC programming concerning the research goals is required e.g.: standard implementation of a new PLC program or implementation of error handling. For both model-driven oo and the state of the art programming approach, HTAs concerning these scenarios have to be conducted. The HTAs provide structured images of PLC programming tasks and therefore enables to develop a comprehension of the different steps that have to be completed in each scenario.

In the following, these scenarios can be compared to pinpoint the differences in the implementation of the programming logic. As a consequence, it is possible to select one or more sections of the HTA in order to develop specific tasks for an empirical study, considering both the research goals and the subjects´ previous knowledge.

Additionally, the necessary quantity of precedent training can be predicted, the subjects´ performance can be assessed and compared and emerging problems during the execution of the task can be deduced as will be shown in the following.

On the one hand researchers can set the main focus on one or more sections of the HTA and on the other hand, they can assess which information research subjects need to execute the tasks. Once the decision for a task has been made, a detailed analysis of the task performance using the HTA can be developed. Finally, the extensive analyzing and the developing of a model solution, including operated as well as cognitive steps, allow researchers to deduce necessary knowledge and skills to complete the given tasks. For this the HTA has to cover the aspects of tool handling as well as the technical and programming aspects of PLC programming.

Like Shepherd describes in [23] training decisions are facilitated through HTA in a number of ways. HTA clarifies the requirements for solving an assignment. It enables training content to be specified by enabling the analyst to weigh up non-training options, including the use of job aids. Further he implies that training decisions must be made systematically as the HTA progresses and that the structure of the HTA facilitates the sequencing of training, both where the training sequence is predetermined and where flexibility of sequencing is preferred.

After analyzing the typical scenarios concerning the research goals and deducing the contents and quantity of training required for specific tasks, it is possible to select sections of the analyzed task sets for experimentation. In this selection process the surrounding influences, like available time for the training and experimentation in conjunction with the available subjects and their previous knowledge can be adequately taken into account.

Performance assessment can be supported through comparison of the tasks analyzed in the HTA and the level of task completion by the participant. To calculate the level of completion, the number of solved tasks from the HTA is divided through the number of all necessary tasks from the HTA. For exact assessment detailed versions of the HTAs are necessary in order to deduct distinct criteria if a specific task is solved or not. Based on the level of task completion the performances also can be easily compared.

In the following we will present the results of an exemplary use case for HTA and tool supported experimental design.

4 Results

At the beginning of the experimental design process an adequate task has to be selected and to be accompanied with a concerted training. The following two main objectives are to be achieved: First, even before the execution of the experiment, there must be hints for a better solution with the new approach than with the comparative approach, i.e. the complexity of a task solution with the respective modeling notations has to be estimated. Second, there must be solubility in the available test time – including the necessary training. The method of choice in order to achieve this is to perform a hierarchical task analysis

In order to control the scope of our study, we focused on analyzing the task of implementing the PLC software. This task is supposed to show significant differences when using state of the art (IEC 61131-3) versus model-driven oo-programming.

4.1 Performed HTAs for Typical Programming Scenarios

In order to define possible experimental tasks we conducted HTAs for two typical PLC programming scenarios, following either the state of the art approach (IEC61131-3) or the model-driven oo approach (UML-Code generator [2]). The first scenario to be analyzed was the implementation of a new PLC program.

As can be seen in Fig.2 both approaches require the participants to identify the process factors, to develop necessary functions and procedures, to implement the program and the hardware mapping in order to be able to fulfill the given main task in this scenario. As the first two tasks are practically identical for both approaches they are not further detailed , instead we focused on the remaining two. On the next lower task level several differences between the two approaches become obvious. While the state of the art approach continues directly with the implementation of function blocks (FBs) the model-driven oo approach requires the subjects to implement their class structure before the main program can be realized. Based on these different task sets specific training exercises were deduced as laid out in the following.

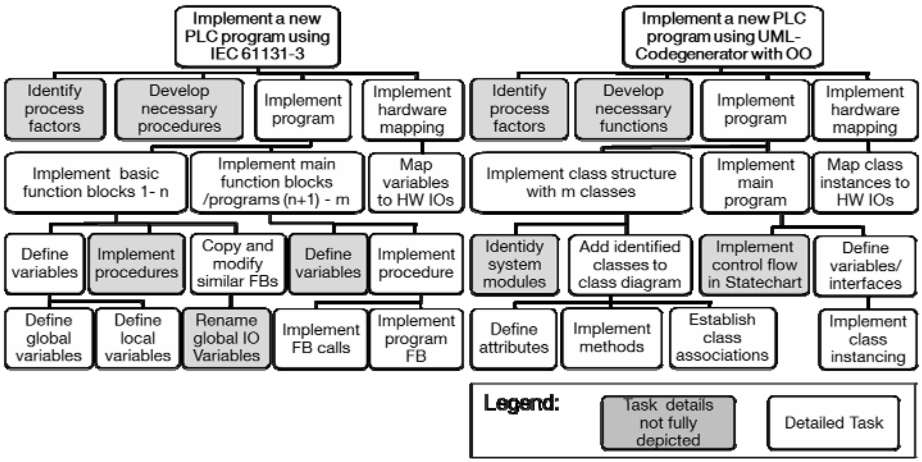


Fig. 2. Abridged graphical HTAs for programming a new PLC program using state of the art PLC programming on the left and the model-driven oo approach on the right.

Subjects using the state of the art approach have to be trained in modular procedure creation, which requires them to understand the system process and to identify similar sequences of functions. In the field of PLC programming this is possible through the "copy and modify" strategy. If the subjects identified the correct procedures, they are able to reuse the implemented logic by copying the whole sequence, e.g. a FB. The copy then has to be modified to react to the right inputs, respectively to control the right outputs. If a variant of the sequence is demanded the copy's logic has to be altered directly. In order to be able to implement and modify the code training in at least one IEC 61131-3 programming language is necessary.

Subjects following the model-driven oo approach need to be trained in modular decomposition of automated systems, so they are able to create feasible class structures as can be seen in Fig. 2. Hence the concept of decomposition has to be taught as well as the use of classes and objects in PLC programming. Structural and functional similarities have to be identified and corresponding classes have to be created using UML class diagrams. Implementation of methods as well as the main program requires training on the usage of State Chart notation for PLC programming and the instancing and usage of objects respectively of their methods.

Finally all subjects need to train the mapping of programmed IOs to hardware IOs.

The research focus in the second scenario, implementation of error handling, was set on the modification of an existing PLC program, shown in Fig. 3. First the subjects have to understand how errors should be handled and develop a specification. The error handling scenario's HTAs show the following tasks of modifying affected modules with greater detail, reaching down to the insertion of states and steps, as these offer the core differences between the state of the art and the model-driven oo approach. The latter requires the subjects to modify all affected classes 1 to m, whereas the state of the art approach demands to modify the FBs 1 to n. For cases in which the number of affected FBs surpasses the number of affected classes, this implies a significant difference. As the training for the first scenario covers all skills necessary to solve these tasks, no additional training is required.

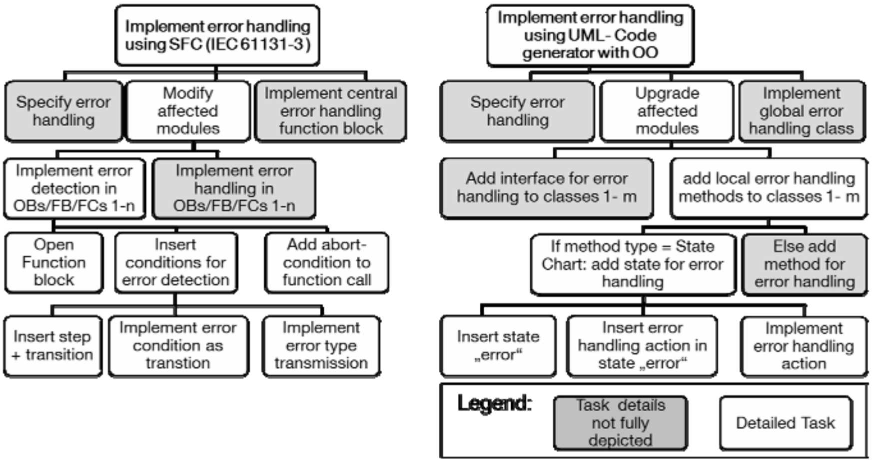


Fig. 3. Abridged graphical HTAs for modifying an existing program to handle errors using state of the art PLC programming on the left and the model-driven oo approach on the right.

For further development of the experiments, the process knowledge of the subjects has to be tested before they enter into the task solving phase. The specific knowledge based on the analyzed tasks has to be ensured, so that the subjects have an understanding of the process and the structure of the given assignment.

4.2 Discussion of the Results

In this paper, enhancements for the usability evaluation of model-driven oo programming approaches in machine and plant automation were described.

It was shown that with sophisticated programming tool support and the use of HTA it is possible to develop an experimental design that meets the requirements concerning research goals as well as training issues. Furthermore, we presented exemplary results of this methodology and potential use cases for the HTA. Based on the presented results, it is necessary to evaluate the new experimental design in the field.

5 Future Prospect

In future studies the described scenarios will be evaluated for different production systems with varying degrees of modularity and similar components. We are confident, that this will support future developments of software engineering environments and model-driven approaches in the field of automation engineering as well as their evaluation process itself. An empirical evaluation of the enhanced experimental design with HTA and tool support will be presented in the future.

References

1. Bonfatti, F., Monari, P., Sampietri, U.: IEC 1131-3 Programming Methodology. CJ International (1997)
2. Witsch, D., Vogel-Heuser, B.: Close integration between UML and IEC 61131-3: New possibilities through object-oriented extensions. In: ETFA 2009, Palma de Mallorca (2009)
3. Vogel-Heuser, B., Wannagat, A.: Modulares Engineering und Wieder-verwendung mit CoDeSys V3. Oldenburg Industrieverlag, Munich (2009)
4. ARC Advisory Group: Programmable logic controllers worldwide outlook. five year market analysis and technology forecast through 2013 (2009)
5. Vogel-Heuser, B., Friedrich, D., Katzke, U., Witsch, D.: Usability and benefits of uml for plant automation - some research results. atp international (2005)
6. Vogel-Heuser, B.: Automation & Embedded Systems Effizienzsteigerung im Engineering. Kassel University Press, Kassel (2009)
7. Friedrich, D., Vogel-Heuser, B.: Evaluating the Benefit of Modelling Notations for PLC-Programming Quality. In: Human Computer Interaction (HCI), Las Vegas (2005)
8. Friedrich, A.: Anwendbarkeit von Methoden und Werkzeugen des konventionellen Softwareengineering zur Modellierung und Programmierung von Steuerungssystemen. Ph.D. thesis, Universität Kassel (2009)
9. Katzke, U., Vogel-Heuser, B.: Design and Application of an Engineering Model for Distributed Process Automation. In: American Control Conference, Portland, pp. 2960–2965 (2005)
10. Katzke, U., Vogel-Heuser, B.: Uml-pa as an engineering model for distributed process automation. In: IFAC World Conference, Prag (2005)
11. Katzke, U.: Spezifikation und Anwendung einer Modellierungssprache für die Automatisierungstechnik auf Basis der Unified Modeling Language (UML), Dissertation Fachbereich 16, University Kassel (2008)
12. Fischer, K., Göhner, P., Gutbrodt, F., Katzke, U., Vogel-Heuser, B.: Conceptual design of an engineering model for product and plant automation. In: Ehrig, H., Damm, W., Desel, J., Große-Rhode, M., Reif, W., Schnieder, E., Westkämper, E. (eds.) INT 2004. LNCS, vol. 3147, pp. 301–321. Springer, Heidelberg (2004)
13. Witsch, D., Ricken, M., Kormann, B.: Plc state charts: An approach to integrate uml state charts in open-loop control engineering. In: IEEE Information Technologies for Sustainable Development (2010)
14. Gemino, A., Wand, Y.: Complexity and clarity in conecptual modeling: comparison of mandatory and optional properties. Data and Knowledge Engineering (2005)
15. Patig, S.: A practical guide to testing the understandability of notations. In: Proc. 5th on Asia-Pacific Conf. on Conceptual Modeling, vol. 79 (2008)
16. Kieras, D.E.: Task analysis and the design of functionality. In: Tucker, A. (ed.) The Computer Science and Engineering Handbook, pp. 1401–1423. CRC Inc., Boca Raton (1997)
17. Annett, J., Duncan, K.D., Stammers, R.B., Gray, M.J.: Task analysis. Her Majesty's Stationery Office, London (1971)
18. Kirwan, B., Ainsworth, L.K.: A Guide to Task Analysis. Taylor and Francis, London (1992)
19. Card, S., Moran, T., Newell, A.: The Psychology of Human-Computer Interaction. Hillsdale, New Jersey (1983)
20. Card, S.K., Moran, T.P.: The keystroke-level model for user performance time with interactive systems. Communications of the ACM 23(7), 396–409 (1980)

21. Lucas, M.R.: Understanding and assessing logic control design methodologies. PhD thesis, Mech. Eng. Dept., Univ. of Michigan, Ann Arbor (2003)
22. Bovair, S., Kieras, D.E., Polson, P.G.: The acquisition and performance of text editing skill: A cognitive complexity analysis. Human-Computer Interaction 5, 1–48 (1990)
23. Shepherd, A.: Hierarchical Task Analysis and Training Decisions. Innovations in Education and Teaching International 22(2), 162–176 (1985)

The Usability Evaluation of Web-Based 3D Medical Image Visualization

Sittapong Settapat[1], Tiranee Achalakul[2], and Michiko Ohkura[3]

[1] Graduate School of Engineering, Shibaura Institute of Technology, Tokyo, Japan
[2] Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
[3] Faculty of Engineering, Shibaura Institute of Technology, Tokyo, Japan
m708501@shibaura-it.ac.jp, tiranee@cpe.kmutt.ac.th, Ohkura@sic.shibaura-it.ac.jp

Abstract. 3D visualization in virtual space simultaneously provides depth information with 2D information visualization ability. Since, web-based e-learning system has become popular alternative framework for improving learning performance and increasing convenience and flexibility to learners. Integrating a 3D medical image visualization into e-learning system aims to accomplish the needs of biomedical engineering education where learners can navigate, browsing, and interact with 3D models of reconstructed medical images. In this paper, we present the usability evaluation results of our web-based 3D medical image visualization comparing with conventional 2D visualization for web-based learning. The experimental results show that 3D visualization method improves learners' education performance with tasks involving 2D information.

Keywords: web3d, 3D visualization, e-learning, distance learning, biomedical engineering education, medical image visualization.

1 Introduction

Since rapid growth of internet environment, web-based e-learning system allows learners, instructors, and facilitators to meet in geographically dispersed places using modern learning pedagogies, methods and facilities [1-5]. Unlike conventional programs in face-to-face teaching and learning, e-learning systems enable anywhere and anytime educations to encourage students to learn either by themselves or from facilitators. However, the available presentation and visualization tools and services mostly are limited in 2D visualization methods including characters, images, animations, and movies. Besides that, immersive virtual environment provides scene simulations and multimedia communication presented in 3D space help influencing learners' behavior and motivation to increase learner satisfaction and efficiency [6-10].

3D virtual environment enrich the learning environment and learning material to be more interactive and perceptive. The integration of 3D virtual environment and web-based e-learning system is not only improves learners' performance and behavior, but this integrated system also provide a convenience and flexibility to learners.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 507–516, 2011.

Meanwhile, in the field of education, information visualization and user interface design affect the effectiveness of learning [6-7]; there are few reports on the effectiveness of the visualization methods for engineering educations such as biomedical engineering education. The biomedical engineering requires the principle understanding in both medical and engineering fields. The understandings in biology and anatomy are basically required, when the medical images are normally used as learning materials.

The results of medical imaging are often given a set of 2D cross-section images, there are several applications that can render these set of image in 3D. However, to visualize these medical images in 3D over the Web require real time 3D reconstruction and visualization services [11]. In general, these services contain high calculation complexity [12]. Putting these services into practical, high computational performance but low cost framework is addressed to promote the web-based 3D medical image visualization service.

Since our research concentrates on a web-based collaborative virtual environment for distance learning [13, 14], we designed and implemented high performance but low cost framework that allows us to reconstruct medical images and to visualize 3D reconstructed models over the Web in real time. Our system implementation was motivated based on a working hypothesis that a 3D visualization method should provide advantages for improving usability more than a 2D visualization method. However in the field of human computer interaction, there are few reports on the usability evaluation of 3D visualization methods for medical images over the Web. Then to validate our hypothesis, we evaluate the usability of our implemented web-based 3D visualization tools and services.

2 Related Research

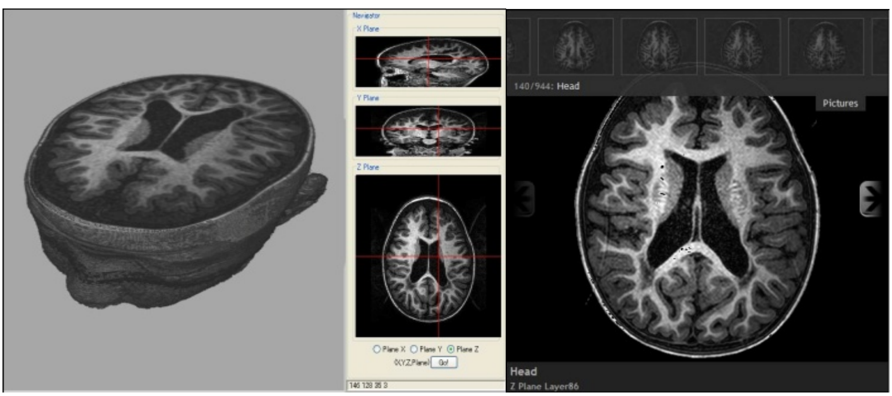
E-learning systems provide the opportunity to learn in both synchronous and asynchronous learning environments [5]. However, to enhance the learning performance, the selection of a certain combination of available e-learning facilities and tools based on instructional models and strategies has major consequences on learner behaviors and performance [2-4].

A collaborative virtual space is an approach for computer-generated scenes and multimedia communication. The generated scenes, which include scene appearance, interaction, navigation, and browsing, affect user feelings and emotions that in turn influence learning behavior and performance [8]. In virtual classrooms, learners interact with classmates and facilitators through avatars that enhance satisfaction and enjoyment [4]. For education, learning usability and effectiveness are as important as learning satisfaction and enjoyment. Enrichment of information visualization for distance learning with better collaboration through web-based e-learning integration improves the learning performance [4, 9]. Moreover, virtual space also offers the ability to spatially present virtual scenes and in 3D interaction [15, 16].

However, the usage of virtual space and 3D visualization is not widespread in engineering education. To promote 3D visualization in biomedical engineering education, we address the integration of 3D visualization and web-based e-learning system. Integrating 3D visualization into web system as a web service, XML based standard file format are recommended. In this paper, we propose the Web3D standard which is the application of XML and VRML technologies Web3D called X3D can deliver interactive 3D objects and 3D worlds across the Internet or CVE. The definition of X3D complies with that defined by the Web3D consortium. Web-based 3D visualization using Web3D standardized visualization tools and a file format increased the maintainability and the flexibility of the system implementation [17]. Web3D allows authors to simply create 3D contents [18]. The created contents can be visualized by any web browser with identical expected visualization using Web3D players. Moreover, Web3D technology allows authors to add more specific interaction for any specific purposes by adding some script and interactive virtual objects. Using this visualization framework, the implemented system becomes a flexible and maintainable scheme that is compatible with web technology.

3 Implemented Application

The reconstructed model from medical images could be visualized over the web using X3D player [19, 20]. However, the X3D player is limited in interaction and operation [21]. Browsing through a reconstructed model normally requires constant page remodeling and reloading, even though we sufficiently minimized the generated model's size to enable loading over web browsers in a reasonable time, so to prevent such an unnecessary model reloading are required. In this paper, we propose our model browsing approach using 2D image navigation. The set of low resolution images is used for the 2D navigation with smaller image resolution.



(a) (b)

Fig. 1. The sample of two visualization methods; (a) Proposed method, (b) 2D slide show

Our 2D navigator shown in Fig. 1a allows users to navigate the model in the 2D mode. The navigator shows the 2D images in three shared planes: X, Y, and Z. In each plane, users can select the position that they want to see, and the image in each plane will be simultaneously changed to the selected coordinates. The model's 2D navigator will not trigger a reload of the model, but the 3D model will reload when users submit the selected position and plane. With this result, users can still seamlessly navigate in 3D through the 2D navigator without unnecessary model reloading. In our system, the medical images can be visualized using a conventional 2D slide show that shown in Fig. 1b.

4 Experimental Research

In this research, we visualized 2D images using two visualization methods; 2D image slide show, and Web3D visualization with 2D navigator. We used a mouse and keyboard for model operation and visualized them via a normal PC display size 36.5 cm. x 23 cm. with screen resolutions 1280 x 960 pixels. The sample of visualization methods are shown in Fig. 1. The motivation behind the system implementation for engineering education was based on working hypothesis that "3D visualization method should provide an advantage for biomedical engineering education with tasks involving 2D information, and students can perform studies using 3D visualization methods effectively."

In order to verify the validity of the hypothesis, we perform an experiment to evaluate our proposed method for biomedical engineering education in anatomy. System usability is defined in ISO 9241 draft standard as the "extend to which a product can be used with effectiveness, efficiency and satisfaction in a specified context of used" [22]. The experiment is intended to compare the level of knowledge obtained through the use of two different visualization systems. Specified tasks are designed to evaluate. After finished the tasks, all subjects need to answer the questionnaire to verify which visualization system is satisfied in anatomy subject for web-based biomedical engineering education.

To carry out the experiment, the three specified tasks were designed to evaluate the level of knowledge obtained through the use of the system.

- In the first task, we introduced subjects to the structures of the human brain through the 2D model with labels, and then subjects have to determine approximate areas of the following structures on the real MR images: Cerebrum, Corpus collosum, Ventricles, Cerebellum, Brain Stem, Frontal Lobe, Left eye, Right Ear and Optic nerve.
- In the second task, subjects have to determine layer positions of the selected MR images on the human head model.
- In the third task, subjects have to determine the name of the structures from the selected human brain areas shown in selected MR image.

The design of these three specified tasks covers the knowledge obtaining from MR image which are "To understand and recognize the human brain structures from image of cross-section model to MR images and vice versa," and "To understand and recognize the approximate real positions on the human head from MR images". To evaluate the system efficiency, spending time of task completion were recorded to compare between two visualization methods. Our two visualization methods, a conventional 2D image slide show method and a proposed 3D visualization method, were performed. Finally, we designed the questionnaire to evaluate the system satisfaction. The Likert 5-scale questionnaires, with the highest score being 5, on seven items shown in Table 1 were used. We group the all items into 3 topics. Items Q1, Q2 are asked to evaluate the information visualization difficulty and satisfaction, items Q3, Q4 are asked to evaluate the usability and operation difficulty, and Q5, Q6 and Q7 are asked to evaluate the overall system satisfaction.

Table 1. The Likert 5-scale questionnaires on 7 items

Item	Question	Item	Question
Q1	It is easy to see the information.	Q5	You are familiar with this system.
Q2	It is easy to understand the information.	Q6	It is user friendly
Q3	The model is easy to control.	Q7	You are satisfied with this system
Q4	The navigator is easy to use.		

4.1 Experimental Procedure

Twelve subjects participated in this experiment, ten graduate students (five males and five females) and two lecturers with MDs (one male and one female). All subjects are Thai nationality. Six of the subjects had experience in interpreting MR images. The other subjects had taken a basic anatomy course. Eleven of twelve subjects used computers daily. Seven of twelve subject used computers daily for educational purposes and the remainder used them several times a week. Six of twelve had experienced virtual space, typically for 3D games and movies; the remainder had never experienced virtual space. Subjects self-reported their familiarity with computers and Internet.

Each subject performed the experiment individually. First, we showed each visualization methods to subjects and train subjects how to use them, and subjects try to use them. All subjects performed the tasks using the same environment. We checked the correctness score that each subject can obtain, and we also recorded the times from start to finish for each visualization method. After the tasks were finished, all subjects answered questionnaires.

5 Experimental Results

For each trail, the collected dependent variables were: time, correctness scores and user ratings of satisfaction and difficulty. Paired samples t-tests were used to find significant contrasts when visualization method effects were discovered. We performed ANOVA to analyze the effect of both gender and the subject's experience in virtual space factors. We chose a significance level of $p<0.05$ for all analyzes. The system efficiency and effectiveness is reported in section 6.1, and the user satisfaction is reported in section 6.2 respectively.

5.1 Efficiency and Effectiveness

First, we present the correctness and time data, and then the statistical evaluation. The results in Table 2 show the raw data of correctness score, the time-to-completion, and the difference between them using 2D and 3D visualizations method.

Table 2. Scores and time-to-complete all tasks of each visualization method

#	Gender	3D Exp.	Score 2D	Score 3D	Paired Diff. (2D-3D)	Time (sec.) 2D	Time (sec.) 3D	Paired Diff. (2D-3D)
S1	Male	Y	62%	94%	-32%	769	665	104
S2	Female	N	75%	98%	-23%	789	704	85
S3	Male	Y	76%	84%	-8%	993	1070	-77
S4	Female	Y	96%	98%	-2%	1170	808	362
S5	Male	N	84%	94%	-10%	1142	959	183
S6	Male	Y	86%	86%	0%	1564	1422	142
S7	Male	Y	100%	100%	0%	1128	869	259
S8	Male	Y	94%	92%	2%	1407	1073	334
S9	Male	N	94%	98%	-4%	855	1586	-731
S10	Female	N	96%	96%	0%	1155	1269	-114
S11	Female	Y	90%	94%	-4%	794	607	187
S12	Female	N	72%	96%	-24%	1300	889	411
Mean			85.4%	94.1%	-8.7% **	1088.8	993.4	95.4
Std. Deviation			11.8	4.8	11.3	257.6	305.4	305.9
Std. Error Mean			3.4	1.4	3.3	74.4	88.1	88.3

1) ** denotes the paired differences is significant ($p<0.05$)

First, we analyzed the system effectiveness using the correctness score. The results show that overall subjects achieved significantly higher scores using the 3D method at $t_{11} = 2.67$; $p = 0.02$. The overall mean for 3D visualization was 94.1% (SD. = 1.4) compared to a mean = 85.4% (SD. = 11.8) for 2D method.

The ANOVA shows that gender and user experience factors did not significantly effect to the correctness score using both 2D and 3D visualization method. For gender factor, The overall mean for male that using 2D visualization was 86.7% (SD. = 13.4) compared to a mean = 84.2% (SD. = 11.1) for female at $F_{1,10} = 0.12$; $p = 0.73$. The overall mean for male that using 3D visualization was 94% (SD. = 4.9) compared to a mean = 94.3% (SD. = 5.27) for female at $F_{1,10} = 0.13$; $p = 0.91$. For user experience factor, the overall mean for experienced subjects that used 2D visualization was 84.2% (SD. = 10.8) compared to a mean = 86.3% (SD. = 13.2) for no experienced subjects at $F_{1,10} = 0.08$; $p = 0.78$. The overall mean for experienced subjects that used

3D visualization was 92.6% (SD. = 5.9) compared to a mean = 96.4% (SD. = 1.7) for no experienced subjects at $F_{1,10} = 1.97$; $p = 0.19$.

We then analyzed the system efficiency using completion time. The results show that overall subjects spending time to finished the task using 2D visualization are not difference by using 3D visualization method at $t_{11} = 1.08$, $p = 0.30$.

The ANOVA shows that gender and user experience factors did not significantly effect to the completion time using both 2D and 3D method. . For gender factor, The overall mean for male that using 2D visualization was 1144.2 sec. (SD. = 306.6) compared to a mean = 1033.5 sec. (SD. = 211.3) for female at $F_{1,10} = 0.53$; $p = 0.48$. The overall mean for male that using 3D visualization was 1095.7 sec. (SD. = 347.3) compared to a mean = 891.2 (SD. = 243.9) for female at $F_{1,10} = 1.39$; $p = 0.27$. For user experience factor, the overall mean for experienced subjects that used 2D visualization was 1117.85 sec. (SD. = 296.54) compared to a mean = 1048.2 sec. (SD. = 216.86) for no experienced subjects at $F_{1,10} = 0.2$; $p = 0.7$. The overall mean for experienced subjects that used 3D visualization was 930.57 sec. (SD. = 281.4) compared to a mean = 1081.4 sec. (SD. = 347.9) for no experienced subjects at $F_{1,10} = 0.69$; $p = 0.42$.

5.2 Difficulty and Satisfaction

The result of user evaluation in system difficulty and satisfaction were shown in Table 3. Paired Samples t-tests were used to find significant contrasts when visualization method effects were founded. The difficulty and satisfaction rating is based on a perceived level of difficulty and satisfaction on a Likert scale of 1 to 5, where 1 was the least and 5 was the most.

In the aspect of information perceiving (Q1), 3D visualization was significant difference from 2D visualization at $t_{11} = -3.37$; $p= 0.01$. From information understanding aspect (Q2), 3D visualization (mean = 4.25) also significant better than 2D visualization at $t_{11} = -3.46$; $p= 0.01$.

For difficulty evaluation, the controllability and usability were asked to evaluate. The paired samples t-test for two navigation method comparison show that there was no significant difference between 2D and 3D navigation and control with mean 3.4, and 3.6 respectively (Q3). However, there was significant for 3D visualization with 2D navigation (mean = 4.2) are more usable than traditional 2D slide show (mean = 3.2) at $t_{11}= -2.87$; $p= 0.02$ (Q4). For system familiarity and user friendly (Q5 and Q6), there were no significant differences between our web-based 3D visualization with 2D navigation and the conventional 2D visualization. Finally, the paired differences between 2D visualization and 3D visualization show that the 3D visualization satisfaction in desire to use (mean = 4.2; mean difference = -1.3) main effect was significant at $t_{11} = -5.2$; $p < 0.001$*.

We asked subjects to evaluate our proposed visualization method for educational purpose. There was significant effect for educational purpose (mean = 4.1) at $F_{1,\,10} = 7.8$; $p = .019$, and eleven of twelve subjects selected our proposed method for web-based biomedical engineering education. From interviewing, there was note that 1 subject selected 2D traditional slide show, because he graduated from Medical Radiation Technology and has a lot of experiences with 2D MR image that agree with our pilot results which evaluated our system by medical radiation technologists. They evaluated our proposed system took an advantage for education. However, most of them still preferred to use traditional 2D slide show which they accustomed with.

Table 3. Subjects' evaluation of system difficulty and satisfaction

	2D Visualization		3D Visualization		Paired Differences (2D-3D)			t	df	Sig.
	Mean	SD.	Mean	SD.	Mean	SD.	Std. Error			
Q1	2.75	1.48	4.08	0.51	-1.33	1.37	0.40	-3.37	11	0.01**
Q2	3.17	0.94	4.25	0.62	-1.08	1.08	0.31	-3.46	11	0.01**
Q3	3.42	1.38	3.58	0.90	-0.17	1.75	0.51	-0.33	11	0.75
Q4	3.17	1.03	4.17	0.72	-1.00	1.21	0.35	-2.87	11	0.02**
Q5	3.58	0.79	3.58	1.00	0.00	1.13	0.33	0.00	11	1.00
Q6	3.25	1.22	3.67	0.89	-0.42	1.44	0.42	-1.00	11	0.34
Q7	2.83	1.03	4.17	0.58	-1.33	0.89	0.26	-5.20	11	0.00**

6 Conclusion

In this study, we implement a prototype application using our proposed framework for biomedical engineering education in anatomy. We designed the experiment to evaluate the effective, efficiency and satisfaction of our implemented system comparing with existing 2D visualization method. Reflecting on the implications of these results, we can answer our original hypothesis and substantiate the following our proposed visualization and navigation method for the Web claims; 3D visualization method provides an advantage for biomedical engineering education with tasks involving 2D information. With integral information visualization, our proposed method offer significant education performance effectiveness and a significant effect satisfaction for educational purpose. Besides that, our visualization approach has no difficulty on controllability and system familiarity compared with the 2D visualization approach. This result conforms to the efficiency evaluation results that there is no difference on learning time spending using both visualization approaches.

The user evaluation results through the use of questionnaire indicate that the advantage of the 3D system is not because of the controllability or familiarity. Adding the third dimension appears to provide more useful information and thus improve the correctness scores. We found that our proposed visualization and navigation method for the Web can improve educational performance. In other words, students can study more efficiently using our proposed framework. Finally, our 3D visualization tool significantly affected the user satisfaction and is proved to be suitable for use as an anatomy learning tool for biomedical engineering students. We believe that our proposed visualization method become more effective contrary with the decrease of number of inexperience virtual space users, or training session should be provided for inexperience users.

References

1. DeRouin, R.E., Fritzsche, B.A., Salas, E.: Optimizing e-learning: Research-based guidelines for learner-controlled training. Hum. Resour. Manage. 2(3), 147–162 (2004)
2. Paquette, G., Ricciardi-Rigault, C., Bourdeau, J., Paquin, C., Liégeois, S.: Modeling a virtual campus environment for interactive distance learning, pp. 17–21 (1995)
3. Paquette, G., Rosca, I.: An ontology-based referencing of actors, operations and resources in eLearning systems. In: 2nd International Workshop on Applications of Semantic Web Technologies for E-Learning. Pays-Bas, Einhoven (2004)
4. Sandanayake, T.C., Madurapperuma, A.P.: Enhancing the Learning Environment by Identifying the Learner Behavior in an e-Learning System. In: Proceedings of the ICEE & ICEER 2009, Korea (2009)
5. Xenos, M., Stavrinoudis, D., Avouris, N., Komis, V., Margaritis, M.: Synchronous collaboration in distance education: a case study on a science course. In: Fourth IEEE International Conference on Advanced Learning Technologies, pp. 500–504 (2004)
6. Hinckley, K., Pausch, R., Goble, J.C., Kassel, N.F.: A survey of design issues in spatial input. In: Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology, pp. 213–222 (1994)
7. Lee, S.H., Harada, A., Stappers, P.J.: Pleasure with products: Design based on Kansei. In: Pleasure with Products: Beyond usability, pp. 219–229 (2002)
8. Liu, Z., Pan, Z.G.: An Emotion Model of 3D Virtual Characters in Intelligent Virtual Environment. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 629–636. Springer, Heidelberg (2005)
9. Schachter, S., Singer, J.: Cognitive, social, and physiological determinants of emotional state. Psychol. Rev. 5, 379–399 (1962)
10. Spillers, F.: Emotion as a cognitive artifact and the design implications for products that are perceived as pleasurable. In: Design and Emotion 2004 (2004)
11. Gering, D.T., Nabavi, A., Kikinis, R., Hata, N., O'Donnell, L.J., Grimson, W.E.L., Jolesz, F.A., Black, P.M., Wells III, W.M.: An integrated visualization system for surgical planning and guidance using image fusion and an open MR. Image 13, 967–975 (2001)
12. Engel, K., Westermann, R., Ertl, T.: Isosurface extraction techniques for web-based volume visualization. In: Proceedings of the Conference on Visualization 1999: celebrating ten years, pp. 139–146 (1999)
13. Settapat, S., Ohkura, M., Achalakul, T.: A Web-based 3D Collaborative Virtual Environment for Distance Learning. In: ICCAS-SICE 2009, p. 128 (2009)
14. Settapat, S., Ohkura, M., Achalakul, T.: A Web-Based 3D Collaborative Virtual Environment for Distance Learning. In: Proceedings of the 13th International Conference on Human-Computer Interaction, pp.465–469 (2009)
15. Zhai, S.: User performance in relation to 3D input device design. ACM Siggraph Computer Graphics 4, 50–54 (1998)
16. Hinckley, K., Pausch, R., Goble, J.C., Kassel, N.F.: A survey of design issues in spatial input. In: Proceedings of the 7th annual ACM Symposium on User Interface Software and Technology, pp. 213–222 (1994)

17. Brutzman, D., Daly, L.: X3D: extensible 3D graphics for Web authors. Morgan Kaufmann Pub., San Francisco (2007)
18. Sankar, G., Rodriguez-Arenas, O.: Web-based Delivery of Scalable 3D Worlds: CMPUT 605 Technical Report (2006)
19. Settapat, S., Achalakul, T., Ohkura, M.: Web-based 3D Visualization and Interaction of Medical Data using Web3D. In: Proceedings of SICE Annual Conference 2010, pp. 2986–2991, SB06.02 (2010)
20. Settapat, S., Archirapatkave, V., Achalakul, T., Ohkura, M.: A Framework of Web3D Based Medical Data Reconstruction and Visualization. In: Proceedings of 12th International Conference on E-Health Networking, Application and Services (IEEE HealthCom 2010), pp. 43–47 (2010)
21. Settapat, S., Archirapatkave, V., Achalakul, T., Ohkura, M.: Web-based 3D Visualization Framework for Medical Data. In: Proceedings of the International Conference on 3D Systems and Applications (3DSA 2010), Tokyo, Japan, pp. 302-305 (2010)
22. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: the CHI is the limit, p. 222 (1999)

A Fitting Software Comparative Usability Study to Investigate Transition Challenges Faced by Hearing-Aid Practitioners

Anil Shankar, Susie Valentine, and Brent Edwards

Starkey Hearing Research Center, Berkeley CA 94704, USA
{anil_shankar,susie_valentine,brent_edwards}@starkey.com
http://starkeyresearch.com

Abstract. Modern application interfaces for desktop PCs such as web-browsers, word processors, and media players share a standardized task oriented user interface (UI). Despite originating from different manufacturers these applications enable users to switch easily between different programs within the same application class, say between web-browsers. When compared to the standardization across these applications there is little or no standardization for different hearing-aid Fitting Software (FS). As a result, practitioners switching between different FS encounter transition challenges while they dispense hearing-aids from different manufacturers. We present usability findings to advocate an enhanced user-centered design process to alleviate the FS transition challenges faced by a practitioner and to improve system-wide usability within a FS. This article presents two main usability findings based on data from twenty-six practitioners who were new to *Inspire*, Starkey's FS; these practitioners were advanced users of FS from three top hearing-aid manufacturers. First, there was significant degradation in task performance for new Inspire users while they performed two standardized tasks. New Inspire users took twice as long to complete these two tasks when compared to an average Inspire user. Second, we found that there were three main categories of usability issues in Inspire; these usability issues coupled with the lack of a standardized UI across different FS exacerbated the transition challenges faced by new Inspire users. Our findings highlight the need for a stronger focus on user-centered design principles for FS manufacturers. We believe that user-centered design is one approach to minimize the effects of competitive marketing and business practices in the hearing-aid industry but still deliver an improved usable system to a practitioner.

Keywords: Fitting Software, Usability.

1 Introduction

Computer user interfaces for complex desktop-oriented software such as web-browsers (Firefox, Safari, Internet Explorer), word processors (Word, Writer), integrated code development editors (Visual Studio .NET, NetBeans, Cocoa), and media players (Winamp, Windows Media Player, Real Player, iTunes) share a standardized task oriented user interface even if they originate from different manufacturers. For example, despite being

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 517–526, 2011.

complex tools, Microsoft's *Word* and OpenOffice's *Writer* present a common user interface for most word processing tasks. These common word processor UIs pose minimal transition challenges (usability issues) for a proficient *Word* user new to *Writer*. In contrast to these applications, hearing-aid *Fitting Software* (FS) UIs lack a standardized user interface primarily due to the competitive business practices in the hearing-aid industry. This lack of standardization across different FS pose significant usability challenges to practitioners (audiologists, dispensers) who are either new to a FS or to those practitioners who regularly use two (or more) types of FS.

Consider such a practitioner, *Jane*. Jane is an expert in one FS, say, FS1 from hearing-aid *manufactuer1*. If Jane wants to hearing-aids from a different manufacturer, say *manufacturer2*, then she either has to attend software training or she has to dedicate her time exploring FS2 and utilizing help sources (training manual, online help, or customer service). The lack of a standardized UI between FS1 and FS2 hinders Jane's ability to perform the *same* task, that is, to dispense hearing-aids. A task-focused user-centered design of the FS from these two hearing-aid manufacturers would result in standardized UI similar to other complex desktop UIs (such as Word and OpenOffice) to ease Jane's task switches between FS1 and FS2.

This article focuses on Starkey's initiative to address the transition challenges faced by new users to our FS, *Inspire*; the next section gives a brief overview of FS for hearing-aids. These users were advanced and expert practitioners who dispensed hearing-aids from the three other manufacturers. We conducted a comparative usability study to identify the most frequent usability stumbling blocks for new Inspire users. To the best of our knowledge, we are not aware of any such comparative usability study of FS to identify transition challenges for new users. Section 3 details our study design, procedure, and tasks. We collected screen recordings, user audio, and self-reports from these three participant groups. This data showed significant degradation in task performance for new Inspire users while they transitioned from their preferred fitting software. We identified the most problematic steps for each participant group using task completion and task success metrics. Our analysis mapped these problematic steps to three categories of usability issues in Inspire: non-standard user interface behavior, lack of awareness of critical features, and navigation issues. Even though we present results related to Inspire's usability, our complementary analysis of Inspire users transitioning to other FS revealed similar classes of usability issues in FS from three other manufacturers.

We organize the rest of this article as follows: the next section briefly introduces Fitting Software (FS) usage in relation to hearing-aids. We give our usability study details in Section 3. Here we outline the study design, the study procedure, and describe the tasks performed by our participants. Section 4 presents task times and task success rate for the three participant groups. Based on these metrics, in Section 5 we identify the main usability issues that adversely affected the task performance of new Inspire users. Finally, we discuss usability initiatives within Starkey to create an iterative, evaluative, user-centered design for developing our FS.

2 Fitting Software for Hearing-Aids

Hearing-aids contain sophisticated digital signal processing algorithms and support many advanced features to improve the quality of life for a hearing-impaired person. A hearing-impaired person either seeks the help of an audiologist or a hearing-aid dispenser to purchase a hearing-aid and to tune the device's parameters to suit the hearing loss. A practitioner (audiologist/dispenser) first connects a hearing-aid to a computer through specialized hardware. Next, she uses sophisticated hearing-aid *Fitting Software* (FS) to adjust the hearing-aid parameters and options compensating for her patient's hearing loss. Note that a FS is specific to a hearing-aid manufacturer. The number of hearing-aid parameters that a practitioner manipulates using the FS is again manufacturer-specific although there are a few basic FS features that are common across various hearing-aid manufacturers. Examples of common adjustments to a hearing-aid include changing the gain provided by the hearing-aid, providing multiple hearing-aid programs and providing feedback control, to name a few. FS plays a very specific and important role in dispensing and fine-tuning hearing-aids.

Inspire OS (Inspire), is the advanced FS developed by Starkey that works exclusively with Starkey hearing-aids. Inspire is a comprehensive software tool that allows a practitioner to adjust and to change a Starkey hearing-aid's parameters and features. A FS from a different hearing-aid manufacturer similarly supports that particular manufacturer's catalog of different hearing-aids. In their daily practice, practitioners dispense hearing-aids from two or more manufacturers. As a result, these practitioners use FS from two (or more) hearing-aid manufacturers to fit hearing-aids.

3 Comparative Usability Study

This section first outlines the main question of our study and then gives details about the two study tasks. Finally, we explain the study design before we describe our results using standard usability metrics in the next section.

3.1 Main Question

We focused on the following question in our comparative study of Inspire versus the fitting software from three other hearing-aid companies:

> What are the transition (usability) challenges that audiologists who are unfamiliar with Inspire face when they perform two standardized fitting tasks?

To answer this question, we designed two standardized fitting tasks that participants performed in four types of fitting software; the next section describes these tasks in more detail. We conducted our study in April of 2009 and participants used FS that was then available from four hearing-aid manufacturers, including Starkey. Our usability analysis in Section 4 focuses on the general transition challenges faced by non-Inspire users. The results are not related to any specific group of participants; hence we do not identify non-Starkey participant groups explicitly but instead represent the groups as P1, P2, and P3, in no particular order. This process allows open disclosure of usability findings while still helping us to use the data for internal bench-marking purposes.

3.2 Two Standard Hearing-Aid Fitting Tasks

We created two tasks related to initial hearing-aid fitting and a follow-up appointment with common hearing-aid complaints. The individual steps across the two tasks were uniform for the four FS except when a FS did not allow a particular step; in this case, we created an equivalent step that would normalize the task performance requirements. In both tasks participants used an actual hearing-aid to simulate the fit. Even though we did not use a hearing-impaired patient to simplify the study procedure we still instructed the participants to perform these two tasks as they would with a patient.

Task 1: Initial Hearing-Aid Fitting. The purpose of this task was to identify the frequent stumbling blocks faced by participants while they performed an initial hearing-aid fitting. We provide the instruction for each step along with rationale to better aid task understanding. The following steps describe the task in more detail:

1. Detect Hearing-Aid: Detect the hearing-aid (using the FS).
2. Quick Connect: Quick connect to the hearing-aid.
3. First Fit: Program hearing-aid for first or best fit according to software. *Rationale*: Hearing-aid manufacturers provide recommended settings for a hearing-aid based on an individual's hearing abilities. When a practitioner first-fits a hearing-aid, she configures the hearing-aid to the manufacturer's recommended settings.
4. Change Target: Change the prescriptive target from the manufacture default to *DSL I/O*. Reprogram the hearing-aid. *Rationale*: Although manufacturers recommend individualized settings for a hearing-aid, there are many commercial fitting formulae available to the practitioner to adjust the hearing-aid settings. In this step a practitioner changed the formula from default to DSL/IO.
5. Feedback Manager: Run the feedback manager. *Rationale*: Running a feedback cancellation algorithm provides a way to reduced or eliminate the feedback, that is, to prevent a patient's hearing-aid from whistling.
6. Setup Programs: Set up hearing-aid programs as follows: (1)Normal, (2)Music, (3)T-coil, and (4)Disabled. *Rationale*: Hearing-impaired people are in many different listening environments in their daily life; sometimes one hearing-aid setting will not work in each environment. Hearing-aids are equipped with multiple hearing-aid programs that can be configured for different listening situations. In this step the practitioner configured the hearing-aid for different environments.
7. Microphone directivity: In program 2 (Music), set the microphone directivity to directional. *Rationale*: A hearing-aid's microphone can either pick up sounds from all directions (omni-directional) or pick up sounds from just one direction (directional). Practitioners have the option to choose a microphone configuration for patient to use.
8. Adjust gain: Patient complains hearing-aids are too loud. In Program 1 (Normal), please reduce the overall gain of the hearing-aid by 4 dB. *Rationale*: Practitioners fine-tune a hearing-aid by adjusting several types of parameters. One such parameter is *gain*, which refers to the amount of amplification provided to the incoming signal.

9. Adjust Compression: Patient also complains that the hearing-aid sound distorted. In Program 1 (Normal), please decrease the compression ratios for the low and mid frequency channels (lower than 2000 Hz) to 1.2 by reducing the gain for soft inputs. *Rationale*: Like gain, compression ratio is another parameter adjusted to fine tune a hearing-aid.
10. Beep Sound: Demonstrate the program change sound (beep) to the patient, via the fitting software. *Rationale*: When a patient changes between two hearing-aid programs the aid plays a signal to alert that patient to the program change. A practitioner uses the FS to demonstrate these program alerts to a patient.
11. Startup delay: Enable a 7-second start-up delay in the hearing-aid. *Rationale*: A start-up delay is used to delay the hearing-aid from reaching a powered on state while the patient is inserting the hearing-aid into the ear.
12. Save & Disconnect: Save changes to a hearing-aid.

Task 2: Follow up appointment with common complaints. We designed the second task to mimic the real world scenario of a follow-up appointment to the initial fit. Similar to the first task, this task's purpose was to identify the transitioning issues for a participant (new Inspire user) while she adjusted the hearing-aid parameters to resolve a common complaint that we included in this task. As in Task 1, participants performed the first two steps: Detect a hearing-aid and Quick-Connect a hearing-aid. Additionally, in this task, participants also adjusted the maximum output of the hearing-aid in this task as described below:

– Adjust max. output: Patient complains that loud sounds are often too loud and sometimes painful. In Program 1 (Normal), please decrease the maximum output of the hearing-aid by 4 dB in the channels above 3000 Hz. *Rationale*: Similar to gain in Task 1, MPO, or maximum power output is a fine-tuning parameter.

3.3 Study Procedure and Design

We first sent out a call for participation using a group mailing list to recruit experienced practitioners (minimum three years of fitting experience) who were familiar with iPFG, Connex, and Genie FS. Next, we selected respondents who had little or no experience with Inspire. With this screening process we recruited 9 each of Genie and Connex experts and 8 iPFG experts.

A participant received $125 for performing the two tasks. We first briefed the participant with a script that outlined the purpose of the study and then gave her an overview of her role in the session. Once a participant confirmed that she was clear about the study procedure she performed the two fitting tasks with Inspire and the FS that corresponded to her expertise. During a session we recorded a participant's screen activity (keyboard and mouse usage) along with audio using TechSmith's Morae Recorder [8]. At the end of the session, a participant filled out our study questionnaire. We made small changes to the Post-Study System Usability Questionnaire (PSSUQ) to suit our study purposes [2]. We collected this post-test expertise to ensure that participants still felt confident about performing the individual task steps and to ensure that any usability issues noted were due to transition challenges.

4 Results

We present results based on the following standard usability metrics; in the next section, we use these metrics to categorize the usability issues that hindered the transitioning for non-Inspire users.

- Average Task-times: With this metric, we compare the time taken by a participant to complete the two tasks with Inspire when compared to their preferred FS.
- Task-Efficiency: We calculated both the time taken on an individual step and also the number of screens accessed to perform the step. To better understand task efficiency we combine the task time and screen accesses to derive an overall efficiency metric.
- Task Success distribution: This metric shows the portion of the task that was completed with ease, the portion of the task completed with difficulty, and the task portion that was completed with error or step failure.

4.1 Average Task Times

Sauro and Lewis have shown that the geometric mean is a better indicator of task-times in usability studies when compared to the arithmetic mean or the median measures [4]. Hence, we use the geometric mean for average task times in this article. Figure 1 shows the task performance of the three participant groups P1, P2, and P3, respectively. In this figure *Inspire* denotes the average time taken by a participant to complete a task within Inspire and *Native* denotes the time taken by the same participant to complete the task using their preferred fitting software (intermediate/expert level of proficiency).

For example, the time on task T1 for the participant group P1 using Inspire was 13.60 minutes when compared to 5.24 minutes with the P1's preferred FS. You can similarly interpret the results for task T2 and for the groups P2 and P3.

Figure 1 shows that there was significant (t-test with 95 percent confidence interval) degradation in task performance when non-Inspire users Inspire to complete the two tasks. Non-Inspire users, on an average, took *twice* as long to complete the tasks when compared to the time they took to complete the same with their preferred FS. To investigate this significant degradation in task performance we next analyzed the efficiency of each step.

4.2 Task Efficiency

We used two measures to calculate the efficiency of each step: the number of screens accessed to perform a step and the time taken to perform the same step. The first measure, screen efficiency, was the ratio of the number of screens accessed by a non-Inspire user to perform a step to the number of screens accessed by an Inspire expert. The second measure, time efficiency, was the ratio of the time taken by a non-Inspire user to the time taken by the Inspire expert group. We combined these two measures to derive the overall task efficiency.

Six Starkey practitioners with advanced (or expert) expertise in Inspire performed the two study tasks. For this group, we calculated the average time taken along with

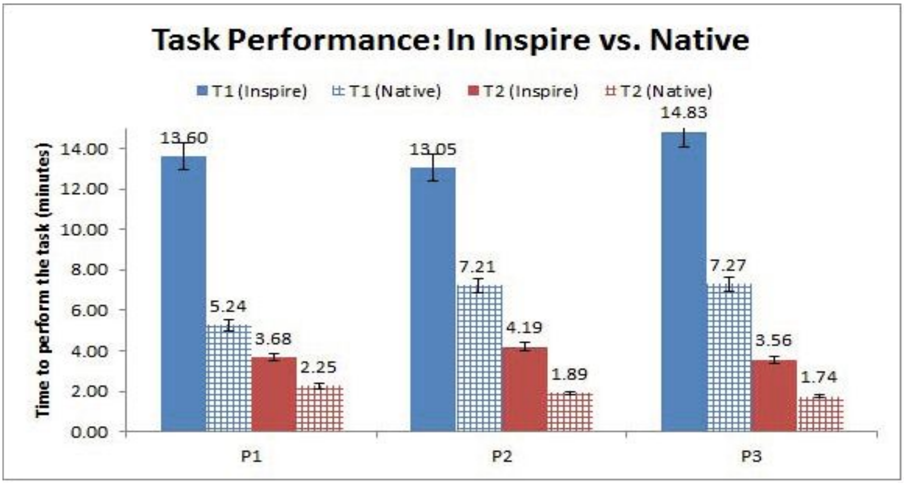


Fig. 1. Task Performance in Inspire when compared to the task performance in the fitting software in which a participant was proficient

the number of screens accessed to complete a step. The overall task efficiency result is relative the performance of these Inspire experts. Since the non-Inspire users were not familiar with Inspire, we considered their performance as first-time use and hence set an acceptance threshold to 70 percent. Figure 2 shows this combined task efficiency.

This figure shows that task efficiency was low for the non-Inspire users when compared to Inspire experts for all the steps except one; only a single step (Setup Programs) met our acceptance criteria.

4.3 Task Success Distribution

To complete our task analysis, we next analysed the task success distribution for each step across the three participant groups. Here, we considered a step as *Completed easily* if the participant completed the step without hesitation or without accessing help. If a participant accessed help, backtracked to previous screens, or corrected an erroneous action then we flagged that step as *Completed with difficulty*. A failure to complete the step was flagged as *Failed to complete*. We completed these assessments with the consensus of an expert in interaction design and of an expert audiologist proficient in all the FS used in this study. Since our participants were mostly first-time Inspire users we considered at least 70 percent of each step *Completed easily* as acceptable instead of using a 100 percent task success rate. For the three participant grops, the following three steps: Set up programs, Quick Connect, and Detect a hearing-aid were acceptable while the remaining of the steps were unacceptable.

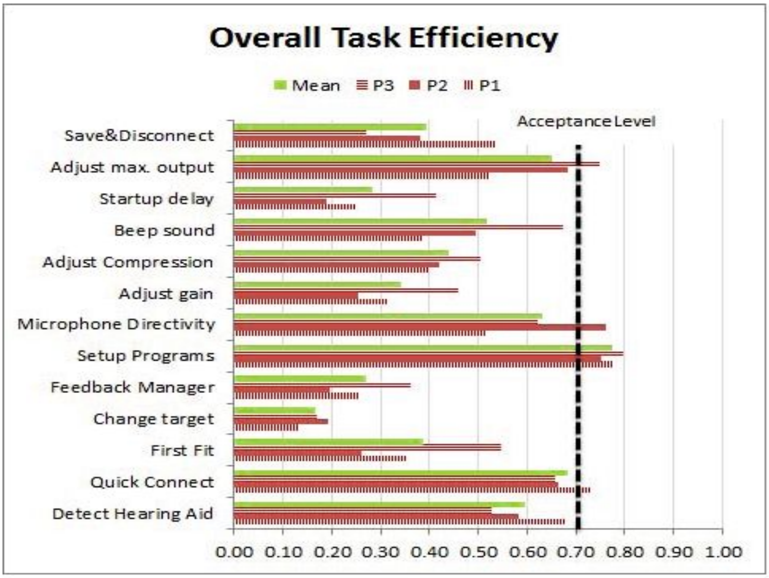


Fig. 2. Combined task efficiency on each step using time and screens accessed across the three participant groups

5 Discussion

While users accomplish tasks with software tools, they create a *mental model* of the software tools (or artifacts) [9]. A user's internal representation of the tool stores structural information about the system along with the functional knowledge about how to perform the tasks. Norman has showed that such mental representations are incomplete, unstable over time, and are parsimonious [3]. When users try to reuse their knowledge of existing systems and if the new system does not support their already existing mental model, task mismatch and keystroke (or mouse) level confusion is the most common manifestation of such user assumptions [10]. The usability issues noted in the previous section were due to such existing mental models of a participant's preferred software when she used Inspire. A combined analysis of these problems demonstrated these three categories of usability issues. Figure 3 shows six such usability problems.

1. Non-standard user interface behavior: A few controls within Inspire did not adhere to a participant's expectation of the interface. For example, in step 3 of Task 2, in Inspire, one click for the decibel adjustment corresponded to a change in $2dB$ while most participants assumed that one click corresponded to $1dB$ change. This behavior contributed to step failure when participants adjusted gain. Another common error was that the *Save* icon did not reflect the saved (or unsaved) state of the session and this resulted in participants clicking this icon multiple times reducing their task efficiency (Figure 3 item-3).
2. Awareness of critical features: Inspire's design adopts a liberal use of tool-tips in the place of labels in different sub-screens. Due to non-standard icons in different

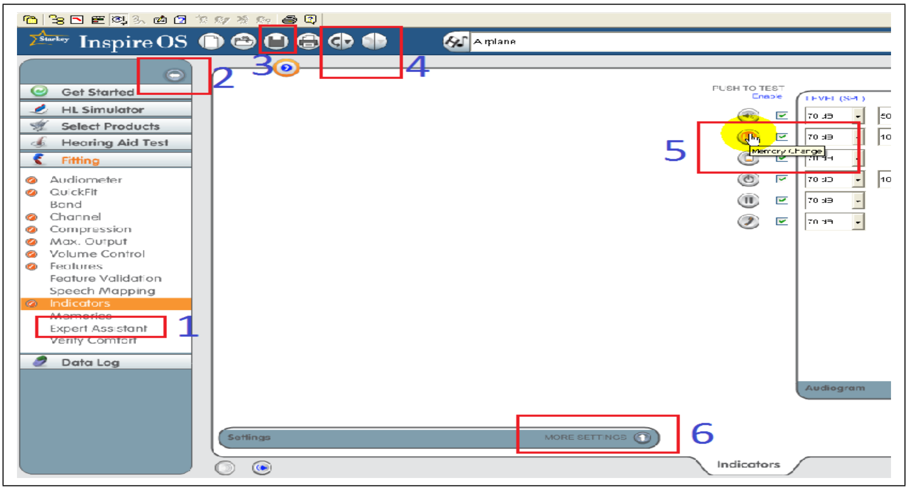


Fig. 3. This figure shows the common usability issues (1) A critical feature, a help wizard, Expert-Assistant hidden in sub-menu. (2) Side-bar that collapses sometimes. (3) Save-icon that does not display status (saved/unsaved). (4) Undo-Redo icons with no labels. (5) Tool-tip that shows the label for an icon. (6) Hidden Settings.

FS, participants *moused* over icons to discover its functionality (Figure 3 item-5). In addition to icon difficulties, Undo-Redo functionality was barely used while a task help wizard was not explored despite being available in the left menu (Figure 3 item-1). We noted similar lack of awareness for changing the prescriptive target (Section 3.2 step-4) and for demonstrating the beep sound (Section 3.2 step-10).

3. Navigation issues: In Figure 3 item-2 shows the left side bar in Inspire that is collapsed in certain sub-screens. This collapsed side bar led new Inspire users astray while they were exploring the interface to accomplish task substeps. Participants were either unable to backtrack to a previous screen to finish a step in a different subscreen or spent time figuring out the current status of the task progression.

These three categories of usability issues significantly hindered the transitioning of users who were unfamiliar with Inspire when they performed two standardized fitting tasks. We were thus able to answer our main study question related to usability challenges for new Inspire users posed in Section 3.1. These common usability issues also demonstrated that Inspire's overall usability could greatly benefit from adopting widely used UI guidelines. Shneiderman proposes one such set of guidelines for user interface design and development. These guidelines address UI consistency, providing user-support, generating appropriate system feedback, and maintaining the locus of control with the user [5,6].

6 Conclusion and Future Work

We highlighted the hearing-aid industry's business practices and marketing issues that have resulted in complex fitting software (FS) that do not share a standardized user interface when compared to widely used desktop applications. Our novel comparative usability study showed that such non-standardized UIs pose significant transitioning challenges to hearing-aid practitioners who were experts in one type of FS but were unfamiliar with another FS. Specifically, we investigated the usability issues faced by three groups of experts switching to Inspire, Starkey's FS. The data from this comparative study showed that there was significant degradation in task performance, task efficiency, and task success for these three groups while they performed two standardized FS tasks. Our analysis identified four main categories of usability issues that led to this task degradation for new Inspire users. We briefly discussed adopting well-established UI guidelines to ameliorate the usability problems faced by these non-Inspire users.

For Starkey, this comparative usability study has provided fresh impetus to create a more user-centered design and development process. We are taking steps to not only address the transition issues faced by new users but to also improve system wide usability. For the hearing-aid industry, in general, we hope that manufacturers will chart progress on user-centered design initiatives and gauge the level of usability acceptance within organizations. One such initiative would be to understand the organizational buy-in using Earthy's usability maturity model that ranks the acceptance levels as follows: skepticism, curiosity, acceptance, and partnership [1,7]. A stronger focus on user-centered design would minimize the effects of competitive marketing and business practices in the hearing-aid industry to ease the transitioning challenges faced by practitioners new to a hearing-aid fitting software.

References

1. Earthy, J.: Usability maturity model: Attitude scale. inuse deliverable d5.1.4s (1998)
2. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq. SIGCHI Bull. 23(1), 78–81 (1991)
3. Norman, D.A.: Some observations on mental models, pp. 241–244 (1987)
4. Sauro, J., Lewis, J.R.: Average task times in usability tests: what to report? In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 2347–2350. ACM, New York (2010)
5. Shneiderman, B., Plaisant, C.: Managing Design Processes. In: Designing the User Interface: Strategies for Effective Human-Computer Interaction, 4th edn., ch. 3, Pearson Addison Wesley (2004)
6. Shneiderman, B., Plaisant, C.: Guidelines, Principles and Theories. In: Designing the User Interface: Strategies for Effective Human-Computer Interaction, 4th edn., ch. 2, pp. 59–106. Pearson Addison Wesley (2004)
7. Stone, D.: User Interface Design and Evaluation. The Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann, New York (2005)
8. TechSmith:Morae Recorder (2010) (accessed 9/23/2010)
9. van der Veer, G.C., del Carmen Puerta Melguizo, M.: Mental models, pp. 52–80 (2003)
10. Waern, Y.: Cognitive Aspects of Computer Supported Tasks. John Wiley & Sons, Inc., New York (1990)

Detection of Software Usability Deficiencies

Dan E. Tamir, Oleg V. Komogortsev, Carl J. Mueller, Divya K.D. Venkata, Gregory R. LaKomski, and Arwa. M. Jamnagarwala

Department of Computer Science
Texas State University
San Marcos, Texas 78666
{dt19,ok11,cm58,dd1290,gl1082,aj21}@txstate.edu

Abstract. In previous research, we have developed an innovative usability evaluation methodology that is based on total effort metrics (TEM). In this paper we present a novel framework for using the TEM approach along with pattern recognition techniques for identifying user interface deficiencies. The methodology, referred to as pinpoint analysis, identifies time segments where the user expends excessive effort in a task-completion interaction-session. The pinpoint analysis methodology provides software engineers with information that can be used for addressing specific usability deficiencies such as poor arrangement of interface elements. The paper describes the new methodology and reports on pattern recognition and TEM based pinpoint analysis tool that is in advanced stages of development.

Keywords: Human Computer Interaction, Usability Evaluation, Pinpoint Analysis, Pattern Recognition, Feature Selection, Principle Component Analysis, Clustering.

1 Introduction

We have developed a novel usability evaluation paradigm that is based on total-effort-metrics (TEM) [1]. The TEM methodology includes logging user manual and eye activities throughout task completion. Logged data is reduced to provide the average value per metric per task. The tasks are derived from one scenario, and referred to as "identical independent" (ID) tasks. For example, the scenario can be "make a travel reservation." Yet, several scenario parameters and constraints such as budget, destination, and hotel amenities vary from task to task. The TEM model, which has been verified through numerous experiments [1], assumes that as subjects are executing ID tasks they become proficient with task execution and their performance improves following a learning curve [1]. Analysis of the average learning curve of subjects enables evaluation of several usability measures such as operability, understandability, and learnability [1].

A recent usability test performed on a software application called Control Studio which is included in Emerson Process Management's DeltaV v10.3 process control system has implemented the TEM framework [1]. Participants completed a series of fifteen ID tasks. Manual interaction activities such as keyboard and mouse clicks were

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 527–536, 2011.

logged and eye movement data was collected with a Tobii x120 eye-tracker operating at 120Hz. Metrics such as time on task, total mouse path traversed, and gazing point location were measured. Using the data collected, we are currently engaged with Emerson Process Management in TEM driven usability evaluation, and in the development and utilization of a pinpoint analysis tool.

In the current research we examine the utility of applying pattern recognition techniques to data acquired from eye tracker and keyboard / mouse activity logging software, in order to identify instances of excessive effort in a user software interaction session and use these instances to pinpoint usability issues. The automatic identification of excessive effort can significantly reduce manual evaluation of software interaction logs. We report on a system constructed to perform the task of pinpoint analysis. Initial results of using the tool and methodology for pinpoint analysis are presented in [2].

2 Background and Literature Survey

2.1 Software Usability

The ISO/IEC 9126 standard defines software usability and identifies Understandability, Learnability, Operability, and Attractiveness as some of the important components of usability [3–4]. Classical methods for usability evaluation can be broadly classified into methods that make use of data gathered from users and methods that rely on usability experts. Usability methods can be further classified into cognitive modeling methods, inspection methods, inquiry methods, prototyping methods, and testing methods [5–10].

Cognitive modeling involves creating a computational model to estimate how long it takes for the subjects to perform a given task. It involves one or more evaluators inspecting a user interface by going through a set of tasks by which understandability and ease of learning are evaluated [5]. Inspection methods differ from cognitive methods; as the experimenters observe users as they use the software. The testing and evaluation of programs is done by designers and HCI experts. This can provide quantitative data as tasks can be timed and recorded. Additionally, qualitative data from users is collected. Although the data collected is subjective, it provides valuable evaluation information [11].

2.2 The TEM Based Usability Model

Several studies indicate that users associate the effort required for accomplishing tasks with the ease of use of the software [1]. Nevertheless, to the best of our knowledge, the current research represents the first attempt to utilize TEM for pinpoint analysis. The research reported in [1], concludes that effort and usability are related. This paper, however, does not address pinpointing issues. In [12,13], the authors use effort metrics to evaluate usability. Their method allows comparison of two or more implementations of the same application, but does not identify the exact location of the problem. In [14] the authors describe the design and test of a defect classification scheme that extracts information from usability problems, but their research does not identify the causes underlying usability problems. In [15], the authors investigate the relations between quantitative data, viewing behavior of users, and web usability evaluation by subjects. They conclude that the moving speed of the gazing points is effective in detecting low usability Web pages. The described research, however, does not attempt to determine and pinpoint the reasons for low usability.

The TEM proposed in [1] provides interface designers and developers with a methodology to evaluate their designs. The approach can be broadly divided into four phases, test design, measurement, analysis, and assessment [1].

In the test design, a scenario is been selected and the ID tasks for this scenario are defined, execution protocols and instructions as well as initial training procedures are determined. During the measurement process, a group of users executes a set of identical independent tasks, which emerge from a single scenario. These tasks differ in key parameters which prevents the users from memorizing the sequence of interaction activities. Throughout the interaction process, certain user activities such as eye movement, keyboard activities, and mouse activities are logged. The analysis phase involves accumulating data for effort metrics that relate to user effort such as the number of saccades, average saccade amplitude, and the number of fixations [1]. Another metric is the time spent on accomplishing each task (ToT). The effort metrics as well as the ToT are compared to a learning curve which reflects user's mastery of software [1]. The final step is the assessment. Using the data obtained in the above steps, the learnability of software systems can be accurately assessed and the point of user's mastery of software can be identified [1]. The same model can be applied to obtain operability and understandability of various systems or different groups of users using the same system. These measures, however, provide general and "low resolution" information to the designers.

Applying the TEM at a higher resolution; that is, measuring user effort expended in relatively short time segments, can provide an accurate means through which an engineer can identify and pinpoint issues in the software or the interface. This process, referred to as pinpoint analysis, can have one of two forms. Inter-pinpoint analysis evaluates the performance of user while executing different tasks trying to identify outlier tasks and check whether the reasons for being outliers can be tracked to a task definition or usability issues. Intra-pinpoint analysis refers to identifying usability issues within tasks. In addition, this analysis also helps graphical user interface designers to make decisions about element placement on displays and determine the level of effort that is related to different widgets[1] [1,12].

2.3 Inter-Pinpoint Analysis

Inter-pinpoint analysis involves identifying tasks that present usability deficiencies and anomalies. For example, a task may present an anomaly because the software enables a special use of the right mouse button, yet some of the users are not aware of this feature [16]. A similar example relates to the fact that some users prefer the toolbar over menus and vice versa.

Inter-pinpoint analysis helps identifying alternative methods to perform a task effectively with less effort. The results of the analysis can be used by the designer to

[1] In this paper, the term widget denotes a basic reusable graphical user interface element such as a scroll-bar.

embed hints for the preferred mode of use in the interface software. Other issues, such as consistency and the availability of help facilities can be identified by the inter-pinpoint analysis process. A more detailed method for analyzing tasks and identifying specific issues with the software is intra-pinpoint analysis.

2.4 Intra-Pinpoint Analysis

Intra-pinpoint analysis can be done in a manual way by having designers and HCI experts watch the video recordings and activity logs of the user's interaction obtained from the eye tracking device and activity logging software. This review helps identifying interaction issues and areas where the user has difficulty while performing tasks. For example, the analysis might reveal that most of the users go into a state of confusion in a specific part of a task and are gazing around the screen to identify the right button to hit and the best way to proceed with the task. This can prompt the designers to rank alternative interfaces based on these findings. That is, every interface may contain some confusion, but the designer is looking for the best alternative or remedy such as rearranging the interface at the relevant time snapshot or to use different shapes and colors for widgets in order to assist the user in distinguishing between widgets. Clearly, manual inspection of the eye tracking video and activity logs is tedious and potentially expensive. An alternative is to use a semi-automatic method applying pattern recognition techniques. Although the usability expert still has to evaluate video recording (hence it is referred to as semi-automatic) the pinpoint method eliminates the need for a person to watch the entire eye tracking video / activity logs and isolates segments of the video and interaction records which present interaction issues; thereby cutting down the evaluation cost and time significantly. A brief description of the pattern recognition techniques used in this research is given in the next section.

2.5 Classical Pattern Recognition Techniques

One of the applications of pattern recognition is classification. The classification process assigns input vectors, where vector components represent metrics or measurements, to a given set of classes [17]. Pattern recognition is generally categorized according to the type of learning procedure used for classification. Supervised learning assumes that a set of instances that have been properly labeled by an expert is provided. This data is referred to as the training data. Next, a learning procedure generates a model that attempts to meet two, sometimes conflicting, objectives: Perform as well as possible on the training data, and generalize as well as possible to new data. On the other hand, unsupervised learning utilizes training data that has not been hand-labeled, and attempts to find inherent data patterns that can be used to determine the correct class assignment for new data instances [17].

Classification algorithms depend on several parameters such as the type of output labels (i.e., class assignment) and on the training / learning method. Additionally, the algorithms differ in the way that inference is performed. For example, inference can be based on probability, non-parametric clustering, and fuzzy logic [18,19].

Generally, the objects that are subject to classification, i.e., the patterns, are represented through a set of measurements (i.e., n measurements) or characteristics referred to as features. Hence, the objects can be considered as vectors in an n-dimensional space referred to as the feature space. Feature selection is a technique for selecting a subset of relevant features for building robust learning and inference models [18]. Feature selection algorithms attempt to reduce the dimensionality of the feature space and the complexity of the recognition process by pruning out redundant, correlated, and irrelevant features. Exhaustive search, heuristic search, and principle component analysis (PCA) are commonly used feature selection methods.

In exhaustive search, all the possible subsets of the features are evaluated and the best subset is selected. This method, however, is generally impractical if the number of features in the set is large [19]. Heuristic search refers to selecting a feature subset by making an educated guess and determining if the selection yields good results. If poor results are obtained, the heuristic procedure examines other subsets. Heuristic based feature selection can be a good alternative in cases where an exhaustive search is impractical [19]. PCA is an unsupervised regression procedure that analyses sample data, such as the set of training patterns, in order to identify an orthogonal coordinate transformation that decorrelates the data and "orders" the information (or variance) associated with the data in the principle components in a monotonically decreasing fashion. In general, as a result of the transformation, most of the information associated with the data is concentrated in the first few components of the new space. This enables ignoring components (axes) that do not carry significant information, thereby reducing the dimensionality of the space used for pattern representation and recognition. Each principal component is a linear combination of the original variables. The principal components as a whole form an orthogonal basis for the data space [19].

The distinction between the principle component analysis and feature selection is that following the PCA, the resulting features are of a different sort than the original features; they do not correspond directly to the set of measurements, and may not be easily interpretable, while the features left after feature selection are simply a subset of the original features.

The classification can be applied via different methods including thresholding, discriminate analysis, decision functions, and clustering [17-19]. The thresholding method classifies input data based on a threshold. All values greater than the threshold are put into one group while the input values lesser than the threshold are classified into another group. One problem with the threshold method is that it is limited to one dimensional data. Hence it can only be applied to individual features, or a combination of features, such as linear combinations or specific components of the PCA. Clustering techniques, however, can be used to efficiently classify multidimensional data.

Clustering is a multi-disciplinary, widely-used, unsupervised data classification algorithm. It involves the assignment of a set of patterns into subsets (clusters) so that patterns in the same cluster are similar in some sense [17-19]. There are several clustering algorithms such as the hierarchical, partitional, density based, and subspace clustering algorithms. Partitional clustering involves partitioning of *n* observations (patterns) into *k* clusters [17-19]. The k-means algorithm, used in this research, is a partition algorithm that attempts to minimize the mean square distance between patterns and cluster centers [17-19]. The center is the centroid of all the cluster patterns. Generally, Euclidian distance is used as the distance measure.

Pattern recognition techniques require the definition of patterns. In this research, temporal segments of user activity records serve as the basic patterns. A segment is defined as a fixed time interval (slice) or as an event based time interval such as the time between two consecutive keyboard/mouse clicks. In the current phase, each segment is represented by a set of measurements composed of: 1) segment duration (for event based segmentation), 2) the average fixation duration for all fixations in the segment, 3) the average saccade amplitude for all saccades in the segment, 4) the number of fixations in the segment, 5) the number of saccades in the segment, 6) the standard deviation of the fixation duration, 7) the standard deviation of the saccade amplitude, and 8) the eye path distance traversed. Other metrics such as manual effort metrics, pupil dilation, and excessive search metrics are under investigation and will be added in future phases of this research.

3 The Pinpoint Analysis Process

Figure 1 illustrates the pinpoint analysis process. The first step includes data (user activity logs) collection. Next, the data logged throughout the user interaction session is used for fixed time interval segmentation and event based segmentation where the events are consecutive keyboard/mouse clicks. Following the segmentation, the set of metrics defined above (section 2.3) is obtained by applying a data reduction program to individual segments. The data obtained from the reduction algorithm for each segment is used to generate vectors of measurements (features). The rest of the phases are used to identify excessive effort segments.

Either feature selection or PCA analysis can be performed in order to reduce the dimensionality of the data considered as inputs to the threshold classifier and/or the clustering algorithm. Since the number of features extracted is relatively small, exhaustive feature selection is feasible. The PCA utility enables selecting the first, first and second, as well as first, second, and third, principle components of patterns as input to the clustering. Additionally, the first component of the PCA as well as arbitrary linear combination of features can be used along with thresholding. The thresholding / clustering algorithms partition the user interaction segments into two classes: “low effort segments”, and “excessive effort segments.” A list of excessive effort segments is generated. The video recording of the eye tracker and the manual interaction records that correspond to the excessive effort segment can be further analyzed by designers and HCI experts to identify the root cause of the excessive effort, determine if it is related to an interface deficiency, propose, and implement remedies to the problem.

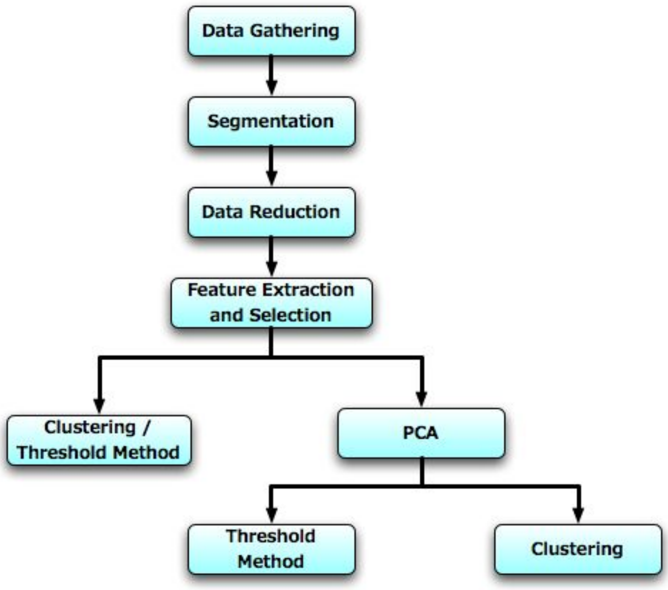


Fig. 1. The Pinpoint Analysis Process

3.1 The Pinpoint Analysis Tool

A pinpoint analysis tool is developed. Figure 2 contains a flowchart depicting the main modules of the tool and their interaction. Figure 3 shows the user interface for the tool. The pinpoint analysis tool has two main phases of operation; calibration and classification. In the calibration phase the tool users select the mode of operation. They can specify parameters such as thresholding versus using clustering, the threshold identification method, the feature selection method, and the feature set. They run the calibration on a subset of the tasks and the tool makes an "Excessive/Non-Excessive" decision on every segment. Furthermore, it enables the users to easily examine the corresponding video recording of the segments and assess the accuracy of the classification using the selected parameters. When the users are satisfied with the accuracy of the tool, they switch to the classification phase where the selected classification method and parameters are applied to different subsets of tasks. The result of the classification is a list of segments identified as excessive effort segments. Again, the tool enables viewing the corresponding video recording. At that point the users are expected to examine the recording in order to identify specific usability issues.

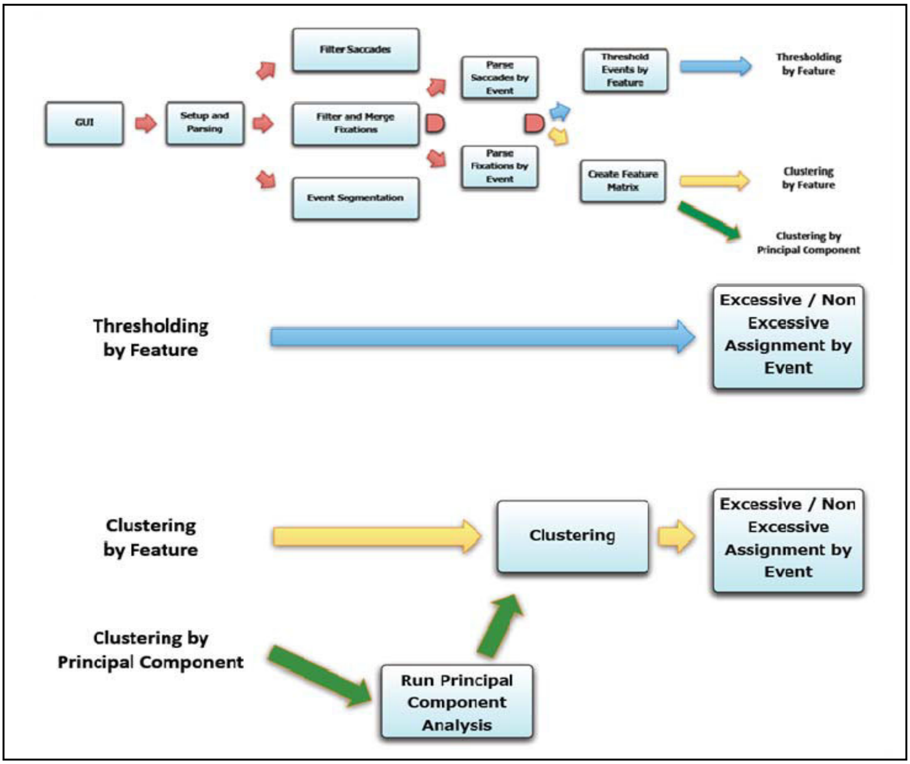


Fig. 2. Pinpoint Analysis Flowchart

4 Experiments and Results

In the calibration mode, the pinpoint analysis methodology can be evaluated in terms of accuracy or classification error. That is, the users can identify errors of type-1 where the pinpoint analysis tool identifies a segment as an excessive effort segment yet the HCI expert does not consider the segment as excessive effort segment. Errors of type-2 where the tool erroneously classifies a segment as non-excessive are identified as well. The total error is the sum of the two types of errors. During the calibration phase the users can bias the tool towards one of the two errors depending on their preference. In the classification phase, the ability of the tool to perform accurate classification as well as the utility of the tool in enabling the identification of user interface deficiencies is evaluated.

The tool reported is in advanced development stage and several experiments measuring the accuracy of the pinpoint analysis methodology have been performed. Some of these experiments, specifically those using the thresholding methods, are reported in [2]. The experiments show that the tool has high potential for reducing the time required for manual inspection of eye tracking video as a means of pinpointing usability issues. Additional experiments using clustering with feature selection and PCA provide similar results.

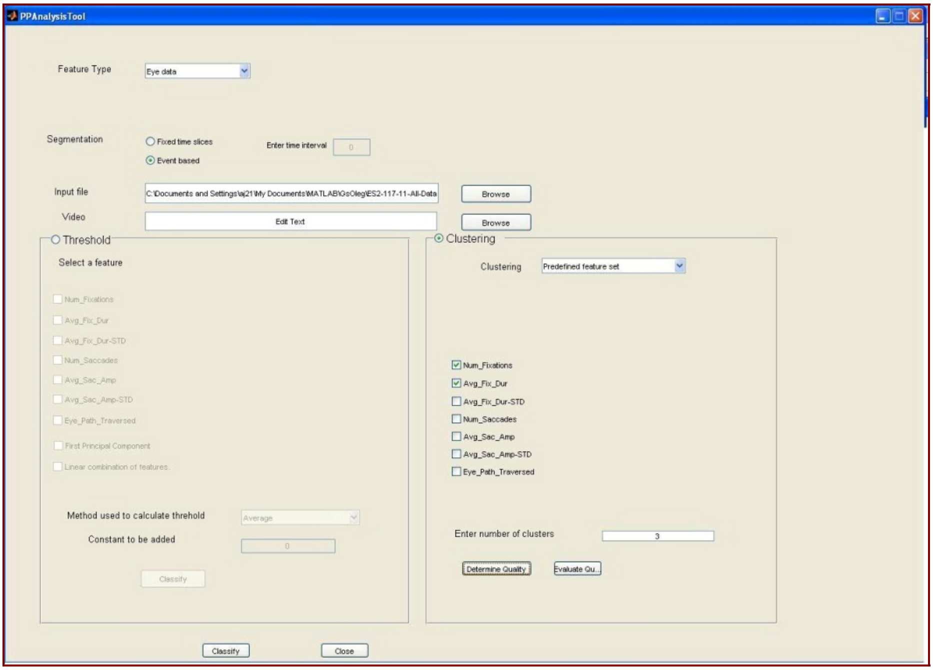


Fig. 3. The Pinpoint Analysis Tool

5 Conclusions

A methodology for pinpoint analysis is described and a tool enabling efficient utilization of the analysis process, which is at advanced stages of development, is introduced. This is a new research and development effort which shows high potential for effective and efficient utilization.

Further work is necessary to finalize the tool development. Furthermore, based on the current experimental results, additional research geared at increasing classification accuracy is due. Future work can include further research related to segmentation methods, evaluation of the utility of additional features, as well as investigation of the performance of the proposed methodology with other types of task assignments and graphical user interfaces.

Acknowledgement. This research was funded in part by Emerson Process Management (www.EmersonProcess.com), an Emerson business.

References

1. Tamir, D.E., Mueller, C.J., Komogortsev, O.V.: A Learning Based Framework for Evaluating Software Usability (2011) (submitted)
2. Komogortsev, O.V., Tamir, D.E., Mueller, C.J., Camou, J., Holland, C.: EMA: Automated Eye-movement-driven Approach for Identification of Usability Issues. In: Marcus, A. (ed.) HCII 2011. LNCS, vol. 6770, pp. 459–468. Springer, Heidelberg (2011)

3. ISO: ISO/IEC 9126-1:2001 Software Engineering-Product Quality-Part 1: Quality Model. International Standards Organization, Switzerland (2001)
4. ISO: ISO/IEC 9126-1:2001 Software Engineering-Product Quality-Part 2: External Metrics. International Standards Organization, Switzerland (2003)
5. Nielsen, J.: Usability Engineering. Academic Press, San Francisco (1993)
6. Barnum, C.M.: Usability Testing Essentials: Ready, Set, Test. Morgan Kaufmann, Burlington (2011)
7. Dumas, J.S., Redish, J.C.: A Practical Guide to Usability Testing. Intellect Books, Portland (1999)
8. Karat, C.M., Campbell, R., Fiegel, T.: Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Monterey, pp. 397–404 (1992)
9. Rubin, J., Chisnell, D.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. Wiley Publishing, Inc., Indianapolis (2008)
10. Tullis, T., Albert, B.: Measuring The User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann, Burlington (2008)
11. Hilbert, D.M., Redmiles, D.F.: Extracting Usability Information from User Interface Events. ACM Comput. Surv. 32, 384–421 (2000)
12. Mueller, C.J., Tamir, D.E., Komogortsev, O.V., Feldman, L.: Using Designer's Effort for User Interface Evaluation. In: 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio (2009)
13. Mueller, C.J., Tamir, D.E., Komogortsev, O.V., Feldman, L.: An Economical Approach to Usability Testing. In: Proceedings of the 33rd Annual IEEE International Computer Software and Applications Conference, Seattle (2009)
14. Noboru, N., Makoto, S., Kazuyuki, S., Ken-ichi, M.: Detecting Low Usability Web Pages Using Quantitative Data of Users' Behavior. In: ICSE 2006: Proceedings of the 28th International Conference on Software Engineering (2006)
15. Vilbergsdóttir, S.G., Hvannberg, T.T., Chong, E.C.: "Classification of Usability Problems (CUP) Scheme: Augmentation and Exploitation. In: Nordic 2006: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles (2006)
16. Tamir, D.E., Mueller, C.J.: Pinpointing Usability Issues Using an Effort Based Framework. In: The IEEE Systems, Man, Cybernetics Conference, Istanbul (2010)
17. Tou, J.T., Gonzalez, R.C.: Pattern Recognition Principles. Addison-Wesley, New York (1974)
18. Tamir, D.E., Kandel, A.: The Pyramid Fuzzy C-means Algorithm. International Journal of Computational Intelligence in Control 2(2) (2010)
19. Duda, R.O., Hart, P.E., Stock, D.: Pattern Classification, 2nd edn. E. Willey International (2001)

Designing Notebook Computers to Ensure a Comfortable User Experience: Effects of Surface Temperature, Material, Locality, and Ambient Temperature

Eric Baugh and Rina Doherty

Intel Corporation
2111 NE 25th Ave,
Hillsboro, OR 97124
United States
{eric.baugh,rina.a.doherty}@intel.com

Abstract. Two studies are described to determine the effect of locality, age, gender, ambient temperature, surface material and surface temperature, on user annoyance during a typing task on notebook computers. The studies were conducted in Oregon and Taiwan, using real computers modified with heaters under the keyboard and palm rests. Computer chassis made from both metal and plastic were studied, and users were exposed to ambient temperatures of both 23 °C and 35 °C. No practically significant effect of locality, age, gender, or ambient temperature was observed, but the ergonomic comfort between metal and plastic surfaces was very different at the same temperature.

Keywords: user experience, computer, temperature, comfort, annoyance.

1 Introduction

High performance mobile computing devices have many engineering challenges, one of which is the thermal design. A proper thermal design not only cools the internal components, but should also provide a safe and comfortable external surface for users. Currently, surface temperature safety limits governing information technology equipment (such as UL 60950 [1]) are very high and generally do not drive the thermal design of notebook computers. A recently published safety standard (IEC 62368 [2]) will impose lower limits, but there is still significant uncertainty as to the usage conditions to which it will apply and whether it will afford a comfortable user experience in and of itself.

Assuming that safety limits are met, the choice of an ergonomic limit involves many trade-offs, and not only along the cost/thermal performance axis. Different market segments, such as value and premium, might require different targets to meet customer expectations, and the same might apply to emerging markets versus developed markets. Notebook computers with metallic construction have a different aesthetic appeal than those constructed from plastic, but this must be balanced against the

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 539–547, 2011.

different manufacturing costs of the two materials. The competitive landscape must also be considered. In order to effectively weigh these trade-offs, data is needed on the subjective impact of surface temperature.

The perception of comfort is subjective, and many factors might influence a user's response to heat. Could using a device in a hotter climate make a user more or less sensitive to the heat of a computer? Does gender or age play a role in heat sensitivity? Does the surface material of the device have an effect? Two studies were carried out to answer these questions. Both studies were designed to determine the influence of age, gender, locality, surface material type (metal and plastic), and ambient temperature (23 °C and 35 °C) on user annoyance of the top (typing) surface temperature of notebook computers.

2 Experimental Hardware

In order to provide a natural usage scenario, real notebook computers were modified with a mix of stock and custom foil heaters applied to the underside of the keyboard tray and the palm rests. An example of a modified system is shown in Figure 1. When an appropriate voltage is applied, the palm rests and the keys of the keyboard are brought to an elevated temperature, which is monitored by a thermocouple embedded in the center of the palm rest or attached to the underside of a key. Figure 2 shows an infrared image of a notebook computer with a plastic chassis where the palm rest heaters can be clearly seen. The heaters underneath the keyboard tray are not well defined because the tray acts as a thermal spreader.

The temperatures for the first study were established using an open-loop calibration procedure. The voltages necessary to attain each desired surface temperature in both ambient room temperatures were pre-determined in a laboratory before the experiment was conducted with the participants. While direct, this method has limitations on accuracy (due to differences between the laboratory conditions and the conditions during the actual experiment) and speed (because it may take a long time for the device to reach the desired temperature when driven at a fixed voltage). In order to overcome these limitations, a closed-loop control was implemented for the second study. This involved a PID (proportional plus integral and derivative) control program running on a master computer which monitored the actual temperature of each thermocouple in the experimental computers and adjusted the voltage in real time. This allowed the surfaces to reach the desired target temperature with a high degree of accuracy (within ±1 °C) in the minimum amount of time since large voltage swings were allowed initially when changing temperatures, and then the magnitude of the voltage changes were reduced as the system stabilized. Even with the closed-loop control system, however, the thermal inertia of the devices meant that about 10-15 minutes was required for stabilization at each new temperature.

In both studies, the voltage was held constant when the participants were using the computers. This replicates the constant power condition that a heavy workload would provide to the typing surface of the computer. Since conduction into the human hand is a more effective means of heat dissipation than the radiation and natural convection modes that a bare surface experiences, the surface temperature of the notebook will drop in the space of a few seconds when touched under normal conditions. The amount of temperature drop on contact is dependent on several things, including the pressure with which the hand presses and the amount of blood flow in the hand. Because there is no standardized way of testing the actual contact temperature, the values in this paper and those in the safety standards cited above represent the initial, pre-contact surface temperature.

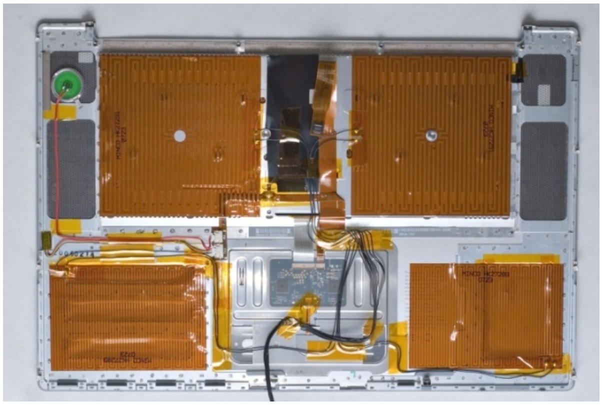


Fig. 1. Foil heaters applied to the underside of the keyboard tray (top row) and palm rests (bottom row) of a notebook computer with metal surface

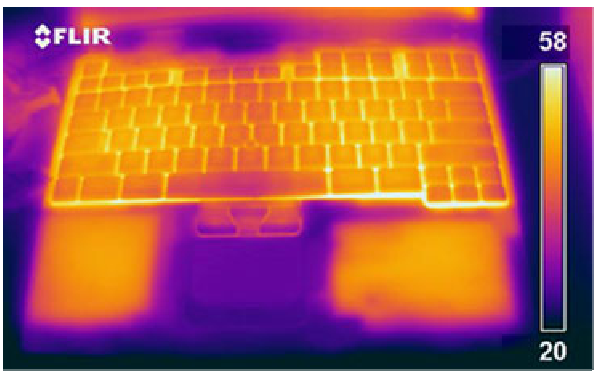


Fig. 2. Infrared image of a notebook computer with plastic surface when the underside heaters are activated. Temperature scale is in degrees Celsius.

3 Experimental Design

Participants between the ages of 18 to 50 years with an equal mix of men and women were recruited from the general public around Portland, Oregon and around Taipei, Taiwan. All participants used computers on a regular basis and did not employ a 'hunt and peck' style of typing (to limit the variability of surface contact during typing).

Since the medical literature indicates that diabetes can cause a loss of heat sensation (diabetic neuropathy [3]), this was also used as a screening criterion.

Tests were conducted in two ambient temperatures, 23 °C and 35 °C. The lower ambient was chosen to be consistent with the nominal temperature used for acoustic testing of information technology products [4] and represents a typical office or home environment in the United States of America. The higher ambient was chosen to represent a thermal stress condition. The studies were conducted in walk-in thermal chambers so that the ambient temperature could be controlled. Because one of the thermal chambers at the Oregon site was rather compact, participants at that site were also screened for claustrophobia. Randomization and counterbalancing techniques were used to reduce order effects for surface temperatures, material types, and ambient temperatures. For the first study, ambient temperature was a between-subjects variable, while for the second study it was a within-subjects variable.

Using a five-point rating scale as seen in Table 1, participants were asked to rate their annoyance level with respect to the warmth they felt from the keyboard and palm rests after two minutes of typing with the computer placed on a table. Annoying was defined as the extent to which the warmth or heat would bother, disturb, or disrupt them while carrying out the typing task. Participants repeated this task on several computers in order to be exposed to the full range of temperatures and material types. For the second study, the participants also changed between thermal chambers with different ambient temperatures.

Participant age, gender, and initial surface temperatures for the two studies are summarized in Tables 2-4. There is a good balance of ages and genders. The temperatures for the metal systems are generally lower than those for the plastic systems, reflecting the difference in expected user comfort given the respective material properties and the results of an informal pilot study. A narrower range of temperatures was used for the second study, based on the results of the first study and compensating for the increase in testing time by having ambient temperature be a within-subjects variable.

Table 1. Participant Rating Scale

Rating	Description	Interpretation
5	Perceptible but not annoying	Some warmth might be noticeable, but not uncomfortable and would continue to work
4	Perceptible and slightly annoying	Heat is noticeable, but not uncomfortable and would continue to work
3	Perceptible and annoying	Uncomfortable, but would continue to work
2	Perceptible and very annoying	Would continue to work only if highly necessary
1	Perceptible and extremely annoying	Intolerable for any length of time and would not continue to work

Table 2. Participant Gender Summary

Study	Total Participants	Male Participants (Portland/Taipei)	Female Participants (Portland/Taipei)
1	81	18/21	20/22
2	45	21/--	24/--

Table 3. Participant Age Range Summary

Study	Total Participants	Age 18-28	Age 29-39	Age 40-50
1	81	24	32	25
2	45	10	20	15

Table 4. Initial Temperature Summary

Study	Surface Material	Ambient (°C)	Initial Temperatures (°C)					
1	Plastic	23/35	41	45	49	53	56	59
	Metal	23	41	44	47	50	---	---
	Metal	35	---	44	47	50	---	---
2	Plastic	23/35	---	---	48	52	56	---
	Metal	23/35	40	44	48	---	---	---

4 Results

This section will review the observed locality, age, gender, material, ambient temperature, and surface temperature effects for user annoyance related to typing on a notebook computer. Participant ratings were averaged to form a mean opinion score (MOS).

4.1 Locality Effect

Only the first study was conducted in two localities, namely Portland having a moderate to cool climate and Taipei having a much warmer climate. Figure 3 shows the analysis of the overall locality effect, indicating that there is no statistically or practically significant effect. The same is true when each material type and ambient temperature is analyzed separately. This result should not be taken to suggest that there are never any geographic differences, but rather that any such differences may have more to do with the experience level of different regions with computers and the nature of the regional marketplace rather than strictly climatic variations.

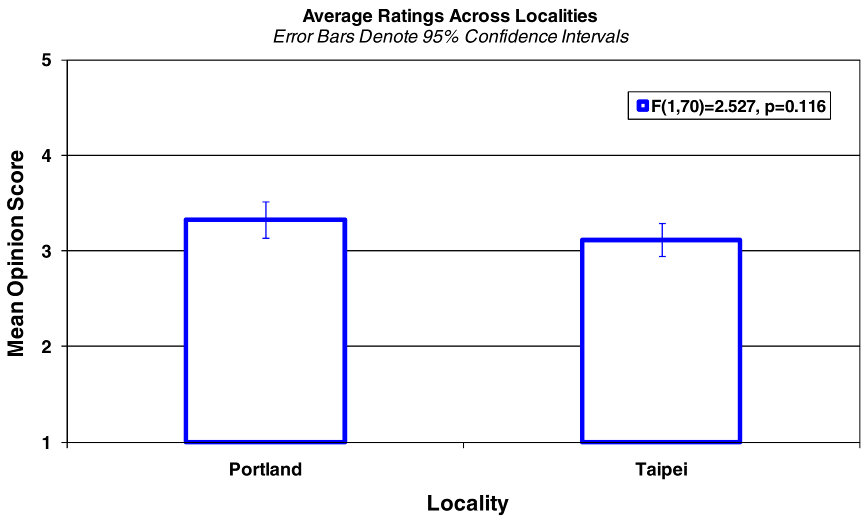


Fig. 3. Overall locality effect

4.2 Age and Gender Effects

Figures 4 and 5 show the overall analysis of the age and gender effects from the first study, indicating that there is no statistically or practically significant effect. The same is generally true when each material type and ambient temperature is analyzed separately, and for the second study as well, with the exception that the gender effect approached the threshold of statistical significance in some cases. However, in all cases the gender effect is small in practical terms, and it is recommended that as long as both men and women are included in a study then the result should be acceptable to either gender.

4.3 Ambient Temperature Effect

Figure 6 shows the analysis of the ambient temperature effect. This is a combination of data from the first and second studies. No temperature dependence is seen in the overall analysis. This is consistent with the fact that burn threshold data is independent of ambient temperature [5]. Medical literature also reports that average skin temperature is relatively insensitive to ambient temperature, changing less than 4.5 °C when the ambient temperature changes by 20 °C [6]. Given these findings, it is recommended that ergonomic limits be specified as fixed temperatures over the range of ambient temperatures studied, rather than as values relative to the ambient temperature.

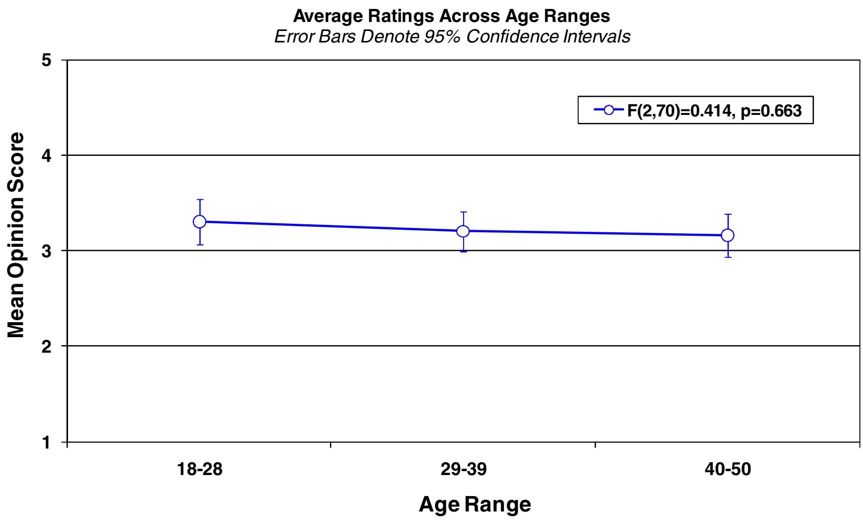


Fig. 4. Overall age effect from first study

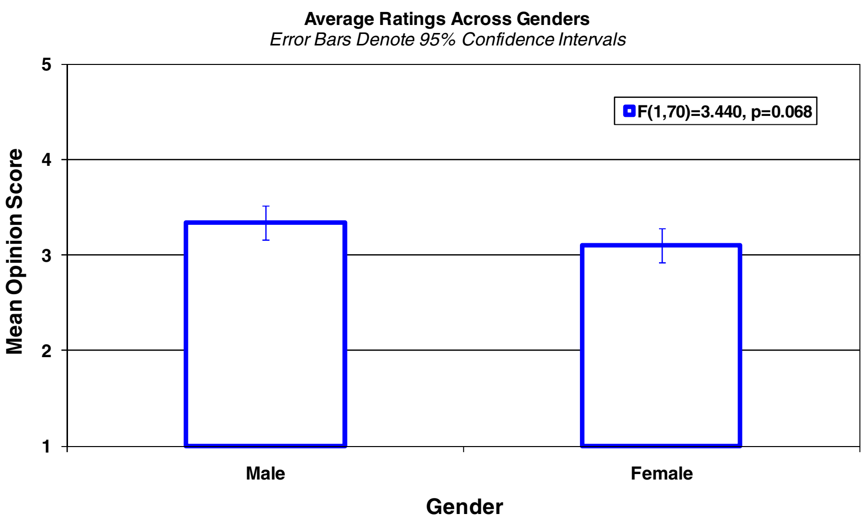


Fig. 5. Overall gender effect from first study

4.4 Material and Surface Temperature Effects

Figure 7 shows the analysis of the effect of the initial surface temperature for metal and plastic surfaces for the second study, which had more accurate starting temperatures due to the use of the closed-loop control system. For the plastic surface, the first and second points are not statistically different, nor are the second and third points. All three points for the metal surface are statistically different from each other. It can be seen that there is at least an 8 °C offset in the temperature between the two surfaces for the same mean opinion score (MOS), with the temperature difference increasing as the annoyance increases (MOS decreases).

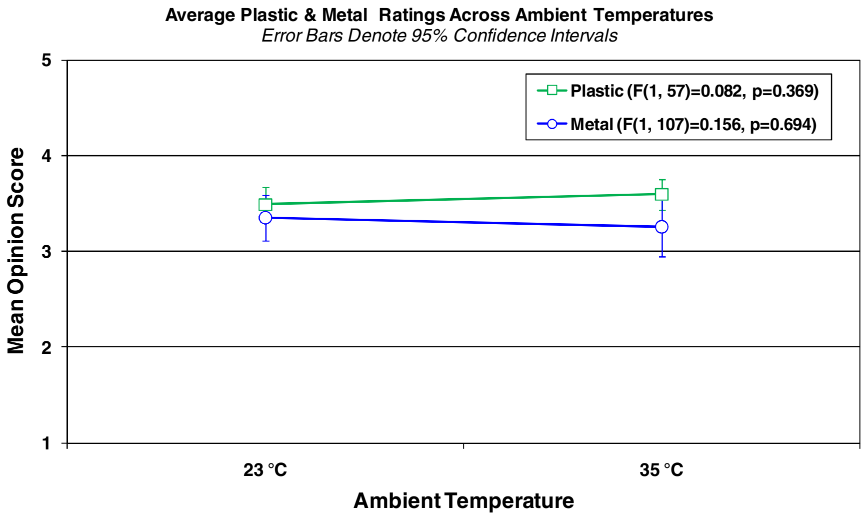


Fig. 6. Ambient temperature effect (selected data from first and second studies)

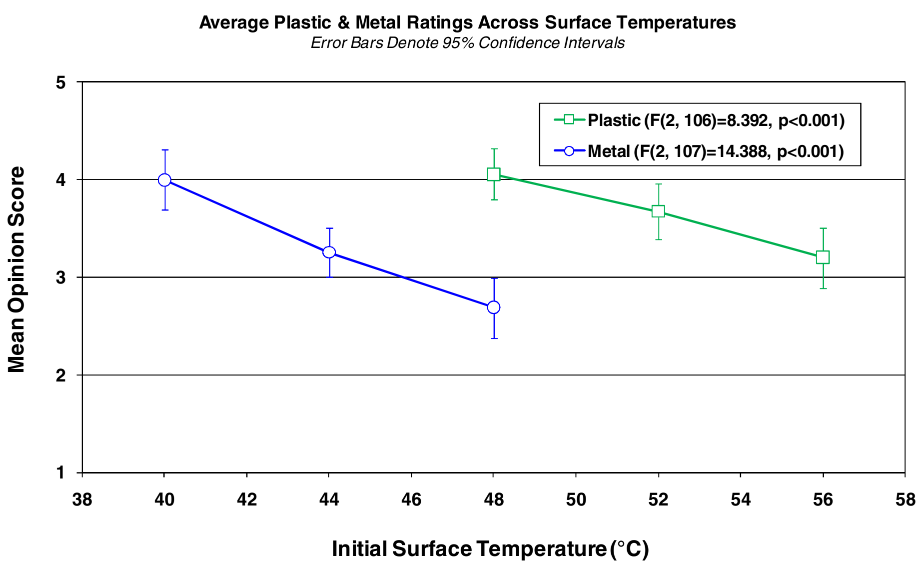


Fig. 7. Effect of initial surface temperature for metal and plastic surfaces (second study)

5 Summary

User comfort and industrial design are both important competitive considerations for notebook computer designers. The findings from these studies clearly show that users tolerate higher temperatures for plastic surfaces than metal surfaces. This must be weighed against the manufacturing costs of each material, the aesthetic appeal, and the cost of the required thermal solution. No dependence in practical terms on age or gender was observed, nor was any difference seen between users in a moderate climate and users in a warmer climate. However, this does not rule out geographic differences based on other factors. No dependence on ambient temperature was observed, and it is recommended that ergonomic limits be specified as fixed temperatures over the range of ambient temperatures studied, rather than as values relative to the ambient temperature. These studies are intended to provide system designers with reliable user perception data to assist them in making informed choices when selecting ergonomic targets for notebook computers.

Acknowledgements. The authors would like to acknowledge the contributions of Eric Salskov, Jessica Tabor, Samy Lin, and Justin Huttula for their help in the execution of these studies.

References

1. UL 60950 Safety of Information Technology Equipment, 3 edn. Underwriters Laboratories, Northbrook, Illinois, USA (2000)
2. IEC 62368-1: Audio/video, information and communication technology equipment - Part 1: Safety requirements, 1 edn. International Electrotechnical Commission, Geneva, Switzerland (2010)
3. Gelber, D.A., et al.: Components of variance for vibratory and thermal threshold testing in normal and diabetic subjects. J. Diabetes Complications 9(3), 170–176 (1995)
4. ECMA-74: Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment, 11th edn. Ecma International, Geneva, Switzerland (2010)
5. ISO 13732-1: Ergonomics of the thermal environment — Methods for the assessment of human responses to contact with surfaces — Part 1: Hot surfaces, 1st edn. International Organization for Standardization, Geneva, Switzerland (2006)
6. Strigo, I.A., et al.: Effect of Ambient Temperature on Human Pain and Temperature Perception. Anesthesiology 92(3), 699–707 (2000)

The Fusing of "Paper-in-Screen": Reducing Mobile Prototyping Artificiality to Increase Emotional Experience

Davide Bolchini and Anthony Faiola

Indiana University – School of Informatics
535 W. Michigan Street, Indianapolis, IN 46202, USA
{dbolchin,faiola}@iupui.edu

Abstract. To address difficult design issues that emerge throughout the prototyping process, interaction designers must rapidly secure user feedback that is both cost-effective and informative during the early stages of the product's life-cycle. To do this, the authors devised a hybrid method of mobile experience prototyping referred to as "paper-in-screen." Ten interaction designers participated in a pilot evaluation study, including a demonstration using the "paper-in-screen" method, hands-on exercise, and a semi-structured interview on the potential and drawbacks of the approach. The study yielded nine themes of qualitative data from which a reflective analysis was performed. Findings suggest considerable support for the method, but also an important contribution to mobile prototyping methodology for interaction designers.

Keywords: Paper prototyping, user experience, design, mobile device.

1 Introduction: Fusing Paper-In-Screen

1.1 Mobile Prototyping Problems and Plus'

The prototype design of mobile applications surrounds issues such as understanding complex software functionality, as well as the integration of customized controls and the interaction modalities of the smaller physical device [1]. Paper prototyping is a common technique for mobile design, but often demands dependence on a facilitator who has a comprehensive understanding of the application and can provide immediate feedback to the test participant [2]. In other words, a full interaction experience is limited. This feedback is in the form of *simulated* interactions that include changing paper-based interfaces or controls within the context of a command sequence, such as clicking/touching buttons or icons. This process also demands a constant reminder to the user about how the various elements of the design fit together in the envisioned functionality of the product; in particular, mobile devices.

Without such human intervention, the participant would not grasp the sophisticated exchange between device functionality, human-mobile interaction (HMI), and system controls. Most importantly, traditional paper prototyping relies on the exclusive use of artifacts that are *detached* from the use of a real, physical device, which generates a

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 548–556, 2011.

sense of artificiality that has not been resolved by design researchers [3][4][5]. Conversely, high-fidelity prototypes that are less rapidly developed provide more actuality in HMI, and more closely resemble the final product.

In sum, standard paper prototypes, though useful and quick to generate, are very limited in delivering the depth of user experience elements that may be necessary to inform design iteration in a more comprehensive fashion. As a consequence, inadequacies of this kind may lead to highly artificial (and ultimately irrelevant) evaluation results. To date, producing high-fidelity electronic prototypes or beta-version releases appear to be the only viable way to have users test or try-out mobile applications *on* their devices. We, however, recommend another intermediate solution.

1.2 A New Agile Prototyping Process: “Paper-in-Screen”

To address difficult design issues and time constraints that emerge throughout the prototyping process, interaction designers must rapidly secure user feedback that is both cost-effective and informative during stages of early design [6][7]. To address this need, the authors have devised a hybrid method of mobile device prototyping referred to as “paper-in-screen,” presented in [8]. Building on this recent research, the authors illustrate here the first systematic evaluation of the technique as continuance of their work.

The “paper-in-screen” method fuses paper and physical device, thereby allowing for a more innovative range of usage paradigms to emerge during prototype generation. The method promises to provide agility, adaptability, and speed in the initial design phase of prototype creation without the need to implement a fully operational high-fidelity prototype. In its conceptual components, “paper-in-screen” taps into a rich research tradition aiming at supporting the design activity with advanced tools and techniques for user interface sketches and prototype designs [9][10][11]. The innovative nuance of the proposed technique rests in its cost-effectiveness, simplicity, and immediate feasibility for the user to anticipate a richer prototype experience during the design process.

From the perspective of mobile application design, “paper-in-screen” greatly enhances the (once limited) quality and delivery of the emotional experience to the user, which is imperative to more adequately inform design iteration [12]. For example, both the appearance of the interface and the physical holding, feeling, manipulating, and touching of the device are significant factors that impact the quality of the user’s cognitive experience. Norman [12] states that understanding the visceral (*physical*), behavioral (*implicit assumptions*), reflective (*aesthetic sensibilities*) levels of the user’s experience is extremely relevant for designing products that better engage the emotion. Moreover, unless we take into account these three factors early on in the life-cycle of device design, in the context of “paper-in-screen,” we may by-pass important elements of the user’s emotional experience.

The “paper-in-screen” method includes six basic steps, ranging from paper prototype design to user experience with a physical device [8]. See Figure 1. No custom software is needed to perform “paper-in-screen,” as every step can be quickly accomplished by using existing practices and common technologies, such as a scanner or digital camera.

"Paper-in-screen" aims at supporting test participants to by-pass the visceral level of initial impact garnered from user interaction with paper substrates. Instead, "paper-in-screen" immediately taps HMI through the user's behavioral experience of look and feel (visual and physical), in which the user can hold, touch, and manipulate the physical device. See Figure 2.

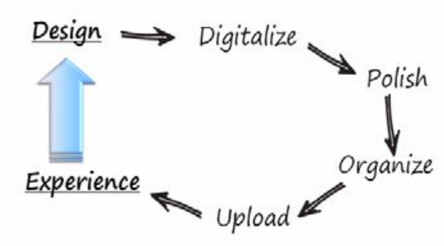


Fig. 1. The "paper-in-screen" lifecycle

Because the iPhone was designed (in part) to both support and amplify the behavioral experience, designing interactive prototypes can impact HMI in ways that are both engaging and intuitive, thereby fully complementing one's implicit assumptions about how a device might behave. As a consequence, "paper-in-screen" may lead to reducing artificiality of the user's experience, resulting in increased design speed, while also avoiding the need to develop high fidelity prototypes.

It is important to note that one significant limitation of "paper-in-screen" (in its current form) is that digitalized paper is not fully interactive, i.e., the only interaction supported is a linear navigation from one screen to another through tapping, sliding, flipping, or using a single button. So far, this is the result of a trade-off between enhancing the emotional user experience and feature complexity, in which the former has been extremely favored to the latter.

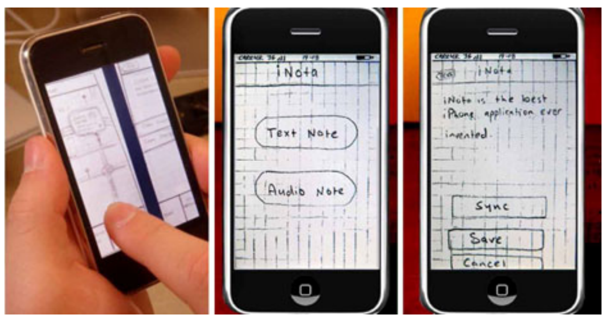


Fig. 2. Interacting with "paper-in-screen" (*left*) and scenario excerpts from the 'paper-in-creen" study (*middle and right*)

More importantly, however, through the use of "paper-and-screen," a technique is now available that affords tactility and amplifies behavior by enabling real interaction experiences. At the same time, interaction designers are more easily positioned to cheaply explore alternative designs based on a truer and more anticipated experience. The expected result of these positive outcomes is the ability to observe and evaluate user behavior through a technique that is extremely malleable, thought provoking, and expressive.

2 Evaluating Paper-In-Screen with Interaction Designers

The goal of this pilot study was to evaluate the effectiveness of "paper-in-screen" among a group of professional interaction designers who work with prototyping on a daily basis.

2.1 Purpose of the Study

Before executing the study, we considered four basic questions from which to measure the impact of concept: (1) Does "paper-in-screen" deliver a *richer emotional experience* for the user by reducing the artificiality of paper prototyping? (2) Does "paper-in-screen" provide an interaction experience that is *more closely aligned* with that of a fully-functional high-fidelity prototype? (3) Does "paper-in-screen" reduce *time and effort* in the early stages of design, while also delivering more useful information about the user experience? (4) Are design researchers able to more readily *observe* user behavior that is more engaged and closer to interacting with a real device?

2.2 Methods and Tools

The evaluation methods of this study included prototype design demonstrations and individual hands-on creation and usage of "paper-in-screen" with 10 professional interaction designers. Participants were recruited within a local, midwestern chapter of the Usability Professionals' Association (UPA) and included designers with solid experience (at least 4-6 years in the field) in mobile application design and techniques of paper prototyping. Participants were introduced to the key elements of "paper-in-screen," then instructed to create their own paper prototypes based on predefined paper templates and design requirements for iPhone applications. After placing the prototypes into the iPhone (Figure 2), usage scenarios were discussed and followed by semi-structured interviews with each participant. Finally, a qualitative analysis of more than 20 hours of audio content was conducted that included transcription and elaboration, extraction of themes, affinity diagramming, and a final evaluation of the data.

2.3 Results

After an affinity diagram was created from the evaluation sessions, the analysis yielded nine themes. We summarize the results of this analysis in the following nine paragraphs, with parallel participant quotes in Table 1.

1. **Perceived Time Efficiency.** Related to *time efficiency*, participants found "paper-in-screen" as a quick way to digitize prototypes that were easy to implement. In sum, participants saw great value in quickly converting their paper into digital form. Participants also found "paper-in-screen" to be time-effective when trying to communicate their mobile design. Others saw the technique as time-effective from a peer-to-peer perspective, i.e., when supporting discussion and critique among interaction designers.
2. **Supporting Sketching and Design Exploration.** Participants concurred that paper prototypes are convenient when quickly testing a design concept in the *early stages*, because they are easy to create, recreate and dispose of. At the same time, they acknowledged that fully-functioning high-fidelity prototypes demand much more time and effort in their creation. Hence, participants believed that "paper-in-screen" was adequate for the early stages of the design process and overall a good way to gather and quickly explore concepts and ideas. Participants highlighted the "sketchy" nature of paper prototypes and how this is also present with "paper-in-screen."

As such, the "digital sketchiness" of "paper-in-screen" was perceived as allowing users to feel more comfortable in expressing their thoughts on unfinished designs, as well as fully understanding the implication of the design on the user's mobile experience. In conclusion, participants felt that "paper-in-screen" made it possible to quickly explore alternative designs by significantly extending the nature of the anticipated experience.

3. **A more realistic mobile experience.** Participants acknowledged that despite the known benefits of paper prototyping, one of its most salient disadvantages is its lack of real-life context. In other words, testing the prototype of an interface of a mobile device is not the same when using a piece of paper as opposed to seeing it on a mobile device. Participants found that "paper-in-screen" helped close the gap between low-fidelity prototypes and a *more realistic user experience* in a more appropriate context [13]. Hence, capturing the realism of the user's mobile experience in context was clearly perceived as a great advantage.
4. **Engaging and Enjoyable for Users.** Broad consensus suggested that paper prototypes fail to *fully engage* users interactively, e.g., when testing paper prototypes, process is bound by the limitations of a piece of paper. Participants concurred that "paper-in-screen" was more *engaging and enjoyable* for consequent users than those made with paper. Participants pointed to the fact that placing such prototypes closer to the intended context of use and "involving more senses" was a key factor. We believe this response was due to the participant's ability to actually handle and touch the device.
5. **Potentially Eliciting Richer Feedback.** Key findings of this study suggested that participants agreed that "paper-in-screen" allowed them the ability and flexibility to gather *better feedback* than paper prototypes. This was because "paper-in-screen" supported their visualization of a broader range of HMI experiences that would otherwise be absent from paper-based prototyping. Participants stated that the "paper-in-screen" experience yielded far more proximity to real-life, which was more likely to enhance user feedback. Regarding the use of iPhone-specific gestures, participants expressed how having the prototype "inside the device" prompted them to engage certain finger gestures and motor skills that would normally not be thought of.
6. **Portability.** As to enhancing the pragmatic aspects of prototype design testing, participants expressed how "paper-in-screen" allowed them to avoid the constraints of a traditional usability lab or a closed-environment. Given the inherent digital quality of "paper-in-screen," several participants expressed how much they enjoyed the ability to "port" the prototypes in different contexts (even outdoors), exchange them electronically, and easily back up and reuse prototypes.
7. **Limited feedback recording and annotation**. Conversely, some participants noted that feedback was more *difficult to acquire* using "paper-in-screen," due to the need for extra equipment to record user feedback, e.g., camera and scanner to digitalize the paper. Participants also mentioned the disadvantage in gathering user feedback because of the perceived learning curve in dealing with the "paper-in-screen" technique. These statements related particularly to the "paper-in-screen" tasks in using the iPhone, e.g., swiping images forward, etc.
8. **Rigidity of the supported interaction**. Although many participants found "paper-in-screen" to be practical in many ways, one voiced concern regarding the

linearity of the scenarios during the test. Another had concerns regarding the inability for the practitioner to easily *make changes "on-the-fly"* to prototypes in the same way paper prototypes provided.
9. **Limited Functionality.** Some participants did not see the benefit of "paper-in-screen" in terms of digitizing a paper prototype. Rather, they held that a more interactive and *fully functional prototype* should be developed if something more than a paper prototype is needed. Even if using formal paper prototypes was not seen as a problem, another participant mentioned how he would adopt "paper-in-screen" if such prototypes could be more interactive.

3 Discussion and Conclusion

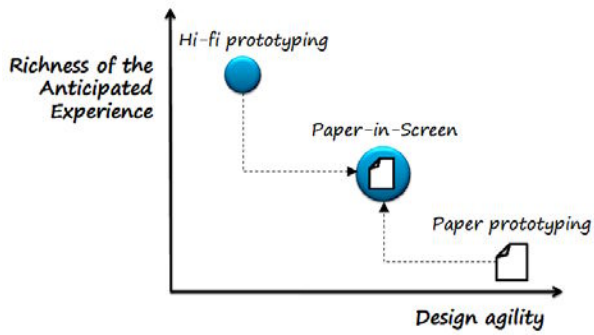


Fig. 3. "Paper-in-screen" slightly compromises prototyping agility to gain a richer anticipation of the mobile experience

In response to the initial research questions, the results clearly indicate the significant potential and transformative value of "paper-in-screen", as well as the limitations and improvements needed. As to the inherent value of "paper-in-screen" for interaction designers, participant feedback focused primarily on the potential of the "anticipated user experience," as noted in Themes 1, 3 and 5. Also, participants agreed that a combination of three enabling factors contributed to achieving greater benefits for both users and designers, i.e., the "sketchy" nature of "paper-in-screen" (Theme 2), the tactile and contextual interaction experience (Theme 3), and the overall innovation of fusing paper and a real device.

A surprising outcome of this study was related to the expected quality of the users' involvement in the prototype evaluation process (Theme 4). Participants suggested that under the normal practice of prototype usability testing (which can often be tedious), "paper-in-screen" offered the possibility of increased user enjoyment and empowerment throughout the process of user involvement. In fact, this factor has the potential to enhance the user sense of ownership and general openness and quality of design input.

The drawbacks of "paper-in-screen" (Themes 7, 8 and 9) point to significant limitations and the need for further research. Considering the trade-offs between supporting an agile design process and the richness of the prototype experience, "paper-in-screen" represents a new compromise between the existing practices (Figure 3).

Table 1. Participant quotes from "Paper-in-Screen" study

Themes	Excerpts from Qualitative Responses
1. Time-Efficient Short-Cut to Rich Prototyping	"…'paper-in-screen' would save a lot of testing time…." "It's pushing as close as possible to the digital realm without all the extra steps to get there". "With "paper-in-screen," you can run the idea through someone quickly and better". "'paper-in-screen' is fast and easy to discuss with your design team".
2. Supporting Early Design	"…'paper-in-screen'" is a bit more flexible than Hi-Fi prototypes and easier to revise". "I like the sketchiness of it, which is important in early designs".
3. Capturing the Realism of the User Experience in Context	"'paper-in-screen' would be hands-on, which is good to gain back the context. A paper sits flat on the table" "While it's not a 100% working prototype, 'paper-in-screen' is closer to the real experience". "The closer you can get to the context, the easier it is to see if you've missed something".
4. Engaging & Enjoyable for the Users	"…paper prototypes might get the same results than with 'paper-in-screen', but the user might be more delighted to participate".
5. Richer Feedback than Paper Prototypes	"…putting the paper prototype on the device is useful especially if there are constraints that are device dependant like text-size". "…['paper-in-screen'] stimulates areas of thought that you normally wouldn't see because it gets you closer to context".
6. Practicality and Portability of the Prototype	"You can't walk alongside someone with a stack of papers and simulate an experience, but you can with 'paper-in-screen'". It would make it easier for me to carry [the prototypes] around. [It makes it] portable". "You could use these images and post them on the Internet for others to test with their iPhones". "With 'paper-in-screen' I could easily control the scenarios because I would not have to worry about sheets of paper getting out of order".
7. Need for extra-equipment to generate prototypes	"A limitation may be that you absolutely need a scanner or some sort of image capturing device" "I would be worried about how to record side notes (meaning, 'on' the prototype)…". "With 'paper-in-screen' users might be inclined to use multi-touch interactions, which could become a problem".
8. Rigidity of Linear Flow and Cost of Changes	"…due to the linear structure of 'paper-in-screen', it would be hard to skip to the home screen if needed, which may take the user out of the 'mindset' a bit". "…I can't use a Post-It on a 'paper-in-screen' prototype to quickly add a pop-up window or correct something".

Table 1. (*Continued*)

9. Limited Functionality	"With functional prototypes, the user would know what the iPhone affords to do in all its features. That becomes a limitation of 'paper-in-screen'".

For example, "paper-in-screen" provides a richer anticipated prototype experience to better identify future quality HMI. At the same time, however, "paper-in-screen" is less flexible than paper-based techniques. Conversely, the technique does not deliver functional artifacts closely resembling the final product as high-fidelity prototypes, but it still enables agile and malleable design exploration early on.

To start addressing these shortcomings, we are exploring strategies to increase the interactivity of "paper-in-screen" by supporting multiple navigation paths and streamlining the process of generating the prototype with ready-to-use online services. We are also systematically comparing (through a controlled study) "paper-in-screen" and paper prototyping to investigate the characteristics in the nature of the actual user feedback elicited by the two approaches.

Acknowledgement. We thank the local chapter of the Usability Professionals Association and all the participants who took part in this evaluation study. We also thank Diego Pulido for the original set up of the "paper-in-screen" technique and his work on data collection and analysis.

References

1. Ballagas, R., Memon, R., Reiners, R., Borchers, J.: iStuff Mobile: prototyping interactions for mobile phones in interactive spaces. In: Proceedings of Pervasive 2006 Workshop on Pervasive Mobile Interaction Devices (2006)
2. Rudd, J., Stern, K., Isensee, S.: Low vs. high-fidelity prototyping debate. ACM Interactions, 76–85 (1996)
3. Krauss, M., Krannich, D.: Ripcord: rapid interface prototyping for cordless devices. In: MobileHCI 2006: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 187–190. ACM Press, New York (2006)
4. Lim, Y., Pangam, A., Periyasami, S., Aneja, S.: Comparative analysis of high- and low fidelity prototypes for more valid usability evaluations of mobile devices. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, Oslo, Norway, pp. 291–300. ACM Press, New York (2006)
5. Lim, Y., Stolterman, E., Tenenberg, J.: The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. Trans. Computer-Human Interaction 15(2), 1–27 (2008)
6. Arnowitz, J.: Effective Prototyping for Software Makers. Morgan Kaufmann, Burlington (2006)
7. Snyder, C.: Paper prototyping: The fast and easy way to design and refine user interfaces. Morgan Kaufmann, Boston (2003)
8. Bolchini, D., Pulido, D., Faiola, A.: Paper-in-Screen Prototyping: An Agile Technique to Anticipate the Mobile Experience. ACM Interactions 16(4), 12–16 (2009)
9. Buxton, W., Lamb, M.R., Sherman, D., Smith, K.C.: Towards a Comprehensive User Interface Management System. Computer Graphics 17(3) (1983)

10. Newman, M.W., Lin, J., Hong, J.I., Landay, J.A.: DENIM: An Informal Web Site Design Tool Inspired by Observations of Practice. Human-Computer Interaction 18, 259–324 (2003)
11. Microsoft Corp. Microsoft Expression Blend 3 Sketchflow, Microsoft Expression at http://www.microsoft.com/expression/products/Blend_Overview.aspx (retrieved July 10, 2010)
12. Norman, D.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books, New York (2003)
13. Beyer, H., Holtzblatt, K.: Contextual Design: Customer Centered Approach to Systems Design. Morgan Kaufmann, San Francisco (1998)

Empathy as Key Factor for Successful Intercultural HCI Design

Rüdiger Heimgärtner[1], Lutz-Wolfgang Tiede[2], and Helmut Windl[2]

[1] Intercultural User Interface Consulting (IUIC)
Lindenstraße 9, 93152 Undorf, Germany
ruediger.heimgaertner@iuic.de
[2] Continental Automotive GmbH
Siemensstraße 12, 93055 Regensburg, Germany
{Lutz-Wolfgang.Tiede,Helmut.Windl}@continental-corporation.com

Abstract. Successful intercultural communication depends on the personal ability to mutually understand the web of belief of the others using empathic capabilities as shown by empirical examples. Only assuming the perspective of a user by the HCI designer to grasp their needs, can lead to good user interfaces of high usability, thereby evoking excellent user experience. Hence, empathy is a key factor for the successful design of intercultural human computer interaction (HCI).

Keywords: Cultural differences, culture, communication, understanding, empathy, intercultural communication, intercultural HCI design.

1 Problems in HCI Design Caused by Cultural Differences

Much cultural background has to be considered when designing the functionality and the interaction for global devices [1]: intercultural HCI design comprises significantly more than merely the implementation of a catalogue of requirements for the user interface like considering different languages, colours or symbols. Successful intercultural HCI design goes far beyond a regular design process by taking into account different mentalities, thought patterns and problem solving strategies that are anchored in culture, for example linear vs. non-linear differences [1, 2]. For example, usage patterns which do not occur in everyday life in the source country can arise in the target country due to different power structures [3], for example flat vs. hierarchical ones. Moreover, the designer must know exactly what the user needs or wants (e.g. why, in which context, etc.) [4]. This knowledge can be determined most precisely by using inquiry approaches or methods based on communication [5]. Just using observation techniques or relying on expert opinions results in less reliable information although these results are very useful, too. However, problems in intercultural communication, particularly those in requirement analysis, inhibit good usability for system design and the related user experience. Therefore, in order to improve cooperation with designers, managers, users and customers, the problems of comprehension must be analysed.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 557–566, 2011.

2 Solving Problems in Intercultural HCI Design by Successful Intercultural Communication Based on Empathy

Several levels of intercultural know-how contribute to successful intercultural HCI design. The communication level constitutes the basic level, followed by the levels of project management, software and usability engineering and HCI design itself on the way to successful intercultural HCI design (cf. Figure 1).

Fig. 1. Levels on the Way to Successful Intercultural HCI Design (Source: IUIC)

Hence, on all levels (strongly influenced by the philosophy of the respective cultures), intercultural communication skills at the basic level can contribute to the solution of problems raised on the upper levels by cultural differences.

2.1 Intercultural Communication Requires Mutual Understanding

For successful (intercultural) usability engineering, an adequate engineering process is necessary to ensure good usability (i.e. when the user understands the developer's device and is thus able to easily operate it satisfactorily), it is necessary that the developer understands the user [2, 6], because they have different points of view [7]. At least the following aspects of the user must be analysed in detail before the product can be developed:

- World view, Weltanschauung (metaphysical approach) of the end-user,
- General knowledge (procedural and factual knowledge) of the end-user,
- The context in which the product will be used by the end-user,
- The tasks the end-user intends to accomplish by using the product.

Only by considering these aspects, intercultural communication as an essential prerequisite for intercultural usability engineering, user interface design, and user experience will it be successful and can it lead to successful international product design.

2.2 Empathy as a Prerequisite for Mutual Understanding

Successful communication depends crucially on the capability for empathy of the people involved [8, 9, 10, 11]. Communication without empathy does not deliver the desired results [12, 13]. This in turn assumes a certain level of confidence and trust (e.g. Principle of Charity, cf. [8]) and includes the knowledge of how to read between the lines of the counterpart's communication depending on culturally different rules. This includes the usage of linguistic rules, for example, Austin's felicity conditions or Gricean maxims [9, 21]. Therefore any literal translation of a conversation is prone to misinterpretation since the extension of the concepts can be different in different cultures ("linguistic relativity", cf. [14]). As context must also be taken into consideration, it is important to consider these aspects in communicating and focus on the intellectual horizon of the communication partner as widely as possible. This can occur in particular through personal and on-site communication and is particularly difficult over the phone due to the absence of mimical and gestical signals. Even more problems arise in intercultural communication compared to intra-cultural communication due to differing world views and the context in which clarification occurs. For this reason, the empathic capability to put oneself in someone else's situation is particularly important. The application of empathy in the end contributes to a successful communication supporting a mutual linguistic code. In particular, intercultural user experience designers must be able to put themselves in the position of the user in order to know and understand his or her intentions and needs, to ideally experience them, and to implement them in the product.

In order to build up not only understanding but also the ability to put oneself in someone else's position, it is initially necessary to be on the same wavelength to find a connection to the other person. This requires the alignment of communication coding (vocabulary and grammar) and to achieve a situation where the other person wants to communicate. Thus a relationship is built up in such a way that future communication remains possible. If this connection is given, it is important to preserve access to the other person's knowledge base ("Web Of Belief", cf. [15]) using a mutual topic of conversation in order to examine the knowledge base of the counterpart in regard to extent, type, and quality. Only then it is possible to find the right "hook" in further conversations and consequently "fetch up" the other person's web of belief at the most relevant point to quickly pick up the same wavelength again. The web of belief contains beliefs and desires derived from premises, assumptions and facts using logical rules and are recursively formed by experience from birth. Through training intercultural competency, approaching the web of belief of a member of other cultures is possible. Thereby, exchange of experiences is very effective, trust can be conferred from one person to another by introducing the persons and critical interaction situations [16] can be weakened.

This works, if it is clear how the other person thinks (i.e. what world view he or she holds i.e. which premises and assumptions about the world he or she has). This is necessary in order to make choices which are relevant to the job at hand and correct for successful communication with a continually expanding set of extra information. This is particularly the case in intercultural contexts. The ability to assess and understand the person's thinking patterns enables an adequate reaction to the people involved. In the same way, the leading and guiding of conversation, e.g., as facilitator or investigator is successfully supported.

2.3 Types of Empathy and Scale of Empathic Sensibility

There are several types of empathy (cf. e.g. [12]):

- Hypothetical empathy by scaling oneself into fantasies to act virtually (e.g. thought experiments),
- Cognitive empathy by changing the thinking perspective by assuming the perspective of another person to see their world with one's eyes,
- Emotional empathy by taking on the concerns of others to sense their feelings,
- Personal distress as the capability of being affected by other human beings in problematic emotional situations or danger,
- Determined empathy as a means to reach a certain purpose (e.g. coping behaviour to become a member of a group),
- Moral empathy applied to manage mutual benefit (e.g. for better understanding or successful communication)

Empathy is related to emotional and social intelligence as well as to tactfulness and altruism.

However, empathy also presupposes the capability to separate oneself from other persons to get the chance to recognize the differences to them and then to put oneself in their position. Within the intercultural context, this requires being aware of one's own cultural standards before it is possible to compare and recognize differences to other cultures [16].

2.4 How to Achieve Empathy?

How can problems in intercultural communication be resolved through empathy? Problems in intercultural communication manifest themselves in critical interaction situations, in which the expectations of at least one party in the interaction are disappointed and can thus result in unease, friction or indeed conflict [16]. Critical interaction situations (CIS) are resolved through cognitive empathy, in other words the ability to see a situation from the other's perspective; to enter the other's cognitive world, by understanding and assuming the other's web of belief for this purpose. How is this possible? The following steps play an important role:

0. If the other person is entirely unknown, initial interpretation is radical: his or her web of belief must be construed almost entirely from scratch [cf. 19].
1. Analyzing the other's output such as body language, actions and statements begins to reveal aspects of the other's web of belief.
1.a. In the absence of external stimuli, this analysis is based largely on observation, for example of behavior (actions), appearance, the effects of actions or behavior and

verbal, nonverbal, and written communication (statements, facial expressions, gestures, documents, etc.).
1.b. In the presence of external stimuli, the analysis may draw on:
1.b.1 questions and an analysis of the output;
1.b.2 an analysis of the output after presenting the counterpart with a specific situation (e.g. a critical interaction situation (CIS)).
The following steps are required after analysis:
2. Generation of a cognitive model (web of belief) of the counterpart
3. Assimilation of the cognitive model of the counterpart
4. Doing step 1 from the counterpart's perspective (in the specific situation)
5. Understanding the counterpart's output (on the basis of the specific situation)
6. Analysis of the specific situation and its root causes
7. Assimilation of the optimal cognitive model by retracing the root causes and retracting the steps which led to the specific situation.
8. Doing step 1 from the counterpart's perspective (in the above specific situation)
9. Comparison of the output with the output in 5.
10. If the output is desirable compared to 5, i.e. positive (no negative actions, friction, etc.), the optimal cognitive model can then be communicated to the counterpart. Steps 1-10 must otherwise be repeated until the desired (positive) output is achieved.

The optimal cognitive model is communicated to the counterpart in the following stages:

a) Assimilation of the cognitive model of the counterpart.
b) Explanation of each step retracted to achieve an optimal cognitive model of the counterpart.
c) Putting the counterpart in one's own place through a) (in the specific situation).
d) The counterpart understands positive output despite the CIS.

If the counterpart is a member of a foreign culture, "Step 5: understanding the counterpart's output (on the basis of the specific situation)" can only be achieved by being trained in intercultural communication (e.g. by workshops, foreign experience, living abroad etc.).

3 Examples for the Necessity of Empathic Capabilities

The following examples refer particularly to experience with projects in the automotive supplier industry concerning China and Germany.

3.1 Greater Project Success by "Guanxi"

The communication between Chinese and Germans is initially so difficult because the rules about how to interpret and understand the utterance of the counterpart, located in the web of belief of the other, are not known. For example, calling a Chinese for support is problematic because the structure, hierarchy and the related rules in China are not known to Germans. The Chinese must strictly follow the rules of the hierarchy in order to support the German's demands and suffers the dilemma of giving support without losing face within his own organization or impeding the good relationship to the one requesting support [18]. In this sense, the Chinese cannot satisfy the request of the German participant without losing face [21]. Therefore, nothing happens. This example will be explained to offer an impression of the difficulty to understand the various effects of the complex relationships in China ("guanxi" [17]). Figure 2 illustrates a part of the structure of a Chinese team hierarchy as well as the relationships between the team members.

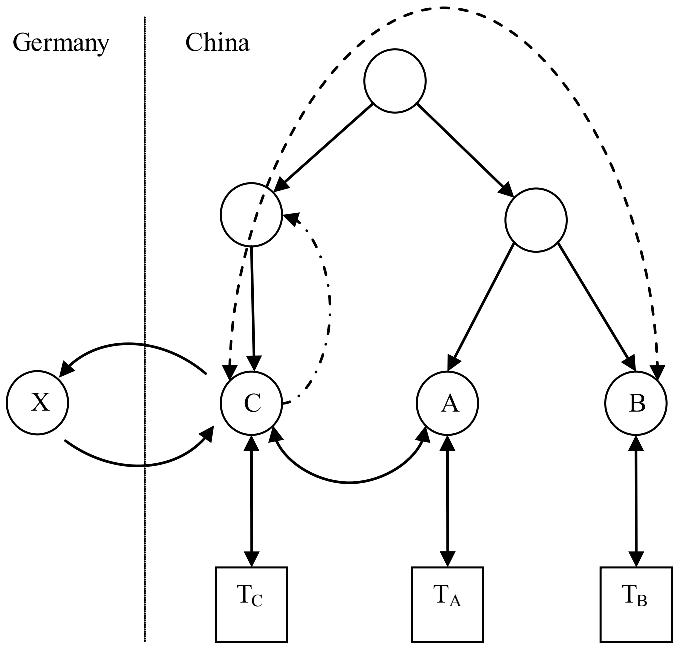


Fig. 2. Chinese Team Hierarchy

Imagine that a German team/project leader X requests support from the Chinese project team T_C. Now, the team/project leader C knows: "Our team T_C need to support X." and communicates to X: "Yes, our team T_C will support you.". However, say to ensure fully the requested support for X, C needs also to ask the team leaders A and B from the teams T_A and T_B providing sub-project support for team T_C. Say C provides support to X as best as he and his team T_C as well as the support from team T_A can provide. However, C does not provide support from team T_B to X. X is not satisfied with the situation and does not understand why he is not supported fully as requested. What is the reason for this undesired situation?

On the one hand, C and A have a good (personal) relationship ("guanxi"), e.g. because they had been at the same university. Therefore, C can ask A for support and bypasses the strict hierarchy rules. On the other hand, C is not connected very well via "guanxi" to B. Therefore, he must go via the official path to B following the whole hierarchy (indicated by the dotted arrow in Figure 2). The hierarchy process is defined by transfer orders “top-down”. By this, C’s responsibility does not include sending requests to his boss or to his boss’s boss because it is not part of his role (represented by the chain-dotted arrow in Figure 2). Hence, he *cannot* do the request according to his job definition and the given responsibility to him. Furthermore, he does not want to “lose face” in front of X by explaining to him his inability to fulfill the task. As a result, the information passed to X is incomplete. Although, X is aware of the requested support/dependencies of the teams T_A and T_B, he cannot start any action like requesting top management support due to missing feedback about why T_A can support and T_B cannot. In principle, the same situation arises in Germany and it is also a known communication issue between German teams – but the details are different. In any case, empathic capabilities are particularly required to put oneself into the situation to be able to master it.

3.2 Better Understanding

Quite often the Germans complain that the Chinese speak "poor English". To understand Chinese people and vice versa it requires a great deal of context. However, German people normally do not know the Chinese context and therefore they do not even understand the questions of the Chinese counterpart. Hence, on the one hand, answers given by Chinese people are often assumed to be incorrect because the context between the explicit statements is not understood by Germans. For example, a Chinese person will always answer in order not to lose face. Thus, a non-Chinese person should not be surprised by answers, which cannot be understood because it seems they do not fit to the question. In this case, one should put oneself in the Chinese’s position and consider why the answer is so unusual as well as analyse, which content needs to be transferred to enable the colleagues in China to understand the question. On the other hand, our direct and open answers probably shock Chinese people, because in their eyes, i.e. from the Chinese perspective, we lose face. However, for Germans, losing face is not as important as for Chinese, which seems to Chinese that Germans lose face without batting an eye. This puzzles Chinese people and alienates them. Empathy is needed.

3.3 Principle of Sustainability

German engineers “like” to improve things. By giving such a request, a Chinese college will not understand why he should develop products of higher quality for "tomorrow" when the customer accepts the quality of the things being sold "today". This phenomenon roots in the Chinese culture: the Chinese are people of today who are typically pragmatic and functionally attuned without strongly focussing on sustainability. Therefore, a good deal of empathy is also desirable here.

3.4 Time-Saving and Problem-Solving by Knowing Relation Networks

Sometimes, projects proceed poorly because one neither knows or shares each other's problems nor is able to solve them. This is because the Chinese network of relations is too difficult to understand in order to always find the right behavioural solutions. Outsiders can, however, certainly get an understandable view of the situation when they have access to very high linguistic and social expertise. In order to be able to understand this network from the inner perspective, however, one needs not only to be born into this perspective, i.e. to be a Chinese, but additionally one needs the understanding by the Chinese that there are complex individual networks and how they can and must be used by the employees. Therefore, in order to understand and consider the importance of this network of relations, people from the target market should be represented in the project team. However, probably even a Chinese cannot get along with the entire Chinese network from an outer perspective.

3.5 Removal of Prejudice for Successful Explanations

The risk of prejudice is very strong when one is in a foreign culture: the impressions are very intensive but also very easily afflicted with prejudices, which, without corresponding reflection, quickly become generalisations. If one has been in a foreign culture for a very long time, i.e. one puts oneself in the position of the other people for a long time, then one's impressions are weaker but one's judgements are more ingrained and finer while taking into account more of the contextual background. Through regular reflection, stereotypes are reduced and attempts to provide explanations for critical interaction situations are more successful. A similar, but much weaker effect can be reached, if there is a chance to establish a multicultural team. The team members can learn from each other if they are facilitated with empathic capabilities (including willingness).

4 Applying Empathic Capabilities in Intercultural HCI Design

Global User Interfaces, which would suit all culture domains, users and contexts, do not yet exist, at least for technical if not for more fundamental reasons. Computers do not yet possess empathy (cf. the so-called „hard problems of AI“ [22, 23]. At the moment computers lack environmental data (through sensors), the complex processing patterns and the respective knowledge of the world needed to develop empathy. Furthermore, the cultural differences involved in the interaction of the user with the system must be integrated in such knowledge of the world so that the system can adjust for it respectively. Finally, even if these challenges were met the so-called 'bootstrapping' problem of adaptive systems would remain. Because the system is not yet acquainted with the user on his first encounter, the system cannot adapt to him. It is a matter of time until the system gets acquainted with the user and can adapt itself to him. At least the following areas must be considered in HCI, i.e. Human-Computer-Interaction: task, context/situation and tools used [24]. In this case the cognitive processes of the user differ from the results of studies or discrete situations due to his cultural and environmental conditioning and personal experience. The concept of the task intended (as well as the task itself) is no longer congruent. That requires the system (computer / machine / tool) to adapt as perfectly as possible to many different aspects, which however has not yet been possible to implement because of the multitude of aspects and the resultant complexity.

As long as the above mentioned problems have not been completely solved, human beings must accordingly attune HCI to the intended cultural domain, user group and context [25]. To do so, the HCI designer must be able to immerse himself in these cultural domains, user groups and contexts in order to extract the relevant requirements for the HCI design.

Hierarchical thought patterns (cf. section 3.1) can exercise an effect on menu structures (i.e. flat vs. deep or highly linked vs. loose) or access rights. Better understanding of the context (cf. section 3.2) leads to better situational adaptation in HCI (i.e. avoiding loss of face through negative feedback in system messages). Understanding networking relations (cf. section 3.4) better facilitates the use of virtual agents (anthropomorphic characters) in Asiatic information systems. Impartiality (cf. section 3.5) in the approach to ‚usability engineering' augments its success and reduces the complexity of the HCI because no detours must be anticipated. By avoiding prejudices unnecessary effort can be circumnavigated (e.g. avoiding the color yellow in China due to association with "red-light-district" connotations). If a situation is known, no prejudices crop up but rather proper evaluations develop, which speeds up the design process and make it specific. That is the way concrete, goal oriented and adaptive HCI design develops and thereby yields highly serviceable and user attuned HCI.

5 Challenges by Empathy

The effort required to train and make use of one's empathic capabilities is a task that is to be taken seriously. How much effort is required over and beyond the use of one's empathetic capabilities? None! As empathic capabilities, however, are not equally distributed nor easily trained in all people, it seems likely that a career as a user experience designer with the goal of achieving successful (intercultural) HCI design is not suited to everyone.

The question of whether the problems in understanding and communicating are more individual empathic problems or based on cultural differences remains open. The question "who gets on well with whom" must still be explicitly explored in relevant studies. It is, however, clear that for communication to be successful, both communication partners in a conversation must be open to each other and in this way be able to use certain basic empathic skills.

6 Conclusion

Empathy is an essential prerequisite for successful intercultural communication, which promotes successful intercultural HCI design and intercultural usability engineering and, as a consequence, good user experience. Hence, it should be ensured and promoted that usability engineers receive, know and apply empathy. This paper just touches the problems, which should be researched in detail in further studies.

References

1. Röse, K.: Methodik zur Gestaltung interkultureller Mensch-Maschine-Systeme in der Produktionstechnik. Univ., Kaiserslautern (2002)
2. Honold, P.: Interkulturelles usability engineering: Eine Untersuchung zu kulturellen Einflüssen auf die Gestaltung und Nutzung technischer Produkte. VDI Verlag, Düsseldorf (2000)

3. Hofstede, G., Hofstede, J.G.: Cultures and Organizations: Software of the Mind, 2nd edn. McGraw-Hill, New York (2005)
4. Holzinger, A.: Usability engineering methods for software developers. Commun. ACM 48(1), 71–74 (2005)
5. Rubin, J., Chisnell, D.: Handbook of Usability Testing. How to Plan, Design, and Conduct Effective Tests. and Conduct Effective Tests. Wiley Publishing Inc., Indianapolis (2008)
6. Nielsen, J.: Usability Engineering. Kaufmann, Amsterdam (2006)
7. Eigenbrode, S., Wulfhorst, J.D., Althoff, D.M., Golberg, C.S., Merrill, K., Morse, W., Nielsen-Pincus, M., Stephens, J., Winowiecki, L., Bosque-Perez, N.A.: Employing Philosophical Dialogue in Collaborative Science. BioScience 57, 55–64 (2007)
8. Dennett, D.: The Intentional Stance. MIT Press, Cambridge (1987)
9. Austin, J.L., von Savigny, E.: Zur Theorie der Sprechakte. Reclam, Stuttgart (2002)
10. Searle, J.R., Kiefer, F., Bierwisch, M.: Speech act theory and pragmatics, pp. 291–312. Reidel, Dordrecht (1980)
11. Schwartz, S.H.: Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology 25, 1–65 (1992)
12. Stueber, K.: Rediscovering Empathy: Agency, Folk Psychology, and the Human Sciences. MIT Press, Cambridge (2006)
13. Stueber, K.: Empathy. Stanford Encyclopedia of Philosophy (2008), http://plato.stanford.edu/entries/empathy
14. Whorf, B.L.: Sprache, Denken, Wirklichkeit. Beiträge zur Metalinguistik und Sprachphilosophie 25. Aufl., Reinbek bei Hamburg (2008)
15. Quine, W.V.: Word and Object. MIT Press, Cambridge (1960)
16. Thomas, A.: Psychologie interkulturellen Handelns. Hogrefe, Göttingen (1996)
17. Budery, R., Huang, G.T.: Guanxi (The art of relationships). Simon & Schuster, New York (2006)
18. Brislin, R.W., Yoshida, T.: Intercultural communication training: An introduction. SAGE Publications, Thousand Oaks (1994)
19. Davidson, D.: Belief and the Basis of Meaning. In: Davidson, D. (ed.) Inquiries into Truth and Interpretation, pp. 141–154. Oxford University Press, Oxford (1984)
20. Grice, P.: Studies in the way of words. Harvard Univ. Press, Cambridge (1993)
21. Victor, D.A.: International business communication. Harper Collings, NY (1998)
22. Görz, G.: Handbuch der Künstlichen Intelligenz. Oldenbourg, München (2003)
23. Winograd, T., Flores, F., et al.: Erkenntnis, Maschinen, Verstehen: Zur Neugestaltung von Computersystemen. Rotbuch Verl., Berlin (1989)
24. Herczeg, M.: Software-Ergonomie: Grundlagen der Mensch-Computer-Kommunikation. Oldenbourg, München (2005)
25. Van Kleek, M., Shrobe, H.E.: A practical activity capture framework for personal, lifetime user modeling. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 298–302. Springer, Heidelberg (2007)

Persuasive Design: It's Not Just about Selling Stuff

Jeff Horvath

Human Factors International,
410 West Lowe, P.O. Box 2020
Fairfield, IA 52556, USA
Jeff.horvath@humanfactors.com

Abstract. When most people think of "persuasive design" in the context of web design, they think about how to persuade a typical online shopper to buy a cool new gadget, a stylish new handbag, or a popular new book. Persuasive design absolutely plays a role in those scenarios, but convincing people to buy something is only one of the places you might employ the principles of persuasion. When I talk to people in government or non-profit about persuasive design, their typical knee-jerk reaction is "that's not for me." When we dig a bit in to their reasoning, it typically comes down to one of two things. They either justify their position by claiming that they are not selling anything, or they take the high moral ground that "persuading" is akin to "tricking" – and, since they are government or non-profit, it's an extra large no-no to trick somebody. In this paper, I will delve deeper in to the how various principles of persuasion can be used for things other than selling. I will generalize the conversation from its more common domain of retail and selling and explain how the ideas behind persuasive design can and do apply to other domains such as government and non-profit. I will provide numerous examples throughout to persuade the reader.

Keywords: persuasion, design, conversion, government, non-profit.

1 Introduction

When most people think of "persuasive design" in the context of web design, they think about how to persuade a typical online shopper to buy a cool new gadget, a stylish new handbag, or a popular new book. Persuasive design absolutely plays a role in those scenarios, but convincing people to buy something is only one of the places you might employ the principles of persuasion.

When I talk to people in government or non-profit about persuasive design, their typical knee-jerk reaction is "that's not for me." When we dig a bit in to their reasoning, it typically comes down to one of two things. They either justify their position by claiming that they are not selling anything, or they take the high moral ground that "persuading" is akin to "tricking" – and, since they are government or non-profit, it's an extra large no-no to trick somebody.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 567–574, 2011.

2 Rethinking Conversion

In some sense, every web site is trying to "sell" something. Some sell things in the traditional sense – you exchange money for a product. Some, however, sell things in a different way – you exchange your time, effort and attention for information, engagement, or a service. A better way of looking at the issue is to think about conversion in a very general sense [1]. Every web site has a goal relative to the site's visitors. If that visitor achieves the goal, they have converted. That might mean purchasing something, but it may also mean finding a piece of information, signing up for some service, or even taking some offline action. In that sense, every web site – including government and non-profit web sites – is trying to persuade its visitors of something. It may be persuading them that the information they find is relevant and trustworthy. It may be persuading them to fill out a census or tax form. It may be persuading them to go see their doctor or sign up for volunteer work. It doesn't have to be about persuading them to buy anything at all.

In their 2008 book *Nudge: Improving Decisions About Health, Wealth, and Happiness* [2], Thaler and Sunstein talk about "nudging" people toward making better decisions. They define a nudge as "any aspect of [a] choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates." They provide a series of examples in numerous domains of designing situations that encourage people to make choices that they may not otherwise make but that are better for them or for society at large. The designers of those contexts – the "choice architects" – nudge them to convert.

Thaler and Sunstein provide a great example of nudging people to do the right thing. The director food services at a large city school system was concerned about the nutritional value of the foods the children were eating and the choices they were making. She could dictate exactly what foods were available and what foods the kids ate, but she knew that providing the kids with choices was important in getting them to eat and be happy about eating at school. So she tried something different. She varied the placement of different foods and the order in which they were displayed. She put different things at eye level. She varied what the kids saw first and what they saw last. In the end, she discovered that she could increase or decrease the consumption of some foods by as much as 25%! By designing the food selection process in different ways, she nudged the kids to eat healthier without restricting their choices or reducing their satisfaction at all. Her goal was to get the kids to make healthier choices. When a kid made a healthier choice, that was a conversion to her. Through some wise design choices, she was able to increase her conversation rate by 25%!

How, then, can a choice architect or a web designer, or a user experience specialist, or anyone affiliated with the design of a web experience… convert/nudge people to do the things they want them to do when visiting a government or non-profit web site? There are a host of different principles [2, 3, 4, 5, 6] that he or she can use to help nudge website visitors. They can speak to some of the visceral cues that speak to our "lizard brain" [3]. They can invoke social cues that make we social creatures want to act. They can make things seem scarce or valuable. They can rely on proven principles of give-and-take or commitment. There are a host of proven principles that can be used. Our friend the director of food services relied on some basic principles about human attention (putting more nutritious things at eye level) and commitment (putting more nutritious things first once children had already chosen them, they would be less likely to undo that choice and select something else later on). The key is that choice architects can design experiences that promote conversion toward goals defined by the site's creators.

Let's look at a few examples.

2.1 Paying Your Taxes

If you don't want people to cheat, don't tell them that lots of other people are cheating. In the mid-90s, the state of Minnesota conducted a study [7] to figure out what might convince citizens to increase their tax compliance – to convert them in to tax payers. The looked at a number of conditions:

- Some people were told that their taxes went to good causes – education, public service, etc.
- Some were threatened with the consequences of not complying – fines, jail time, etc.
- Some were offered help if they needed it in order to submit their taxes.
- Some were told that over 90% of Minnesota tax payers had already submitted their taxes.

Which of these conditions had an impact? Only the last one. The principle that seemed to be at work here was that people were doing what they thought everyone else was doing. When they thought everyone was cheating, they did too. When they found out that most everyone else was doing what they should, they did too. It's a simple nudge, but an effective one. It didn't force users to do anything they didn't want to or make it any harder to file their taxes. It just nudged citizens to convert in to tax payers instead of tax evaders.

Figure 1 shows a few tax-related web sites [8, 9, 10, 11]. Each uses some principles of persuasion – "Free", imagery of "people like me", and others. None of them, however, use the sort of social learning [4] that the MN study shows to help increase conversion of tax compliance.

2.2 Census

The United States government is Constitutionally mandated to take a census of the entire U.S. population every ten years. The data is used for all sorts of secondary purposes, but the primary reason for counting everyone up is for apportionment of seats in the House of Representatives. The process of counting us all up goes something like this:

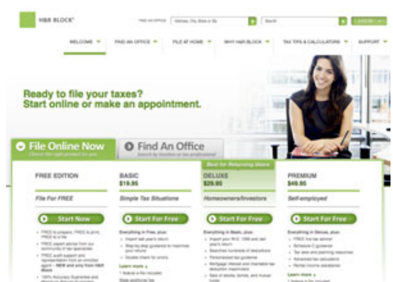
- Send a survey out to every known household
- Census workers walk neighborhoods and collect data on everyone else – people who did not return their surveys, homeless individuals, etc.

While the specific data is not readily available, it is safe to say that it is far faster, easier, and cheaper to get the results that come in via surveys than it is to hire an army of workers to scour the nation. All of this work is paid for by our tax dollars. I for one would like to see my tax dollars spent wisely. I'd like to see as many people as possible fill out the survey. For the 2010 survey, 74% of households filled them out. If I'm the choice architect for the Census, then I would like to find as many ways as possible to get people to fill out the survey. There are plenty of methods and channels to encourage this (tv ads, print mail, social media, etc.), but the web site is one of them. The designers of the web site could and likely did try to use many kinds of nudges. Figure 2 shows one such attempt [12]. It shows the percent of responses from each state and locale. The upper Midwest seems to have done particularly well for some reason.

(a) IRS

(b) WI Department of Revenue

(c) TurboTax



(d) H&R Block

Fig. 1. Tax-related Web Sites

Every design can always be improved and the Census's design could likely have used some additional principles of persuasive design, but this particular one is largely the same as that discovered by the Minnesota tax collection group – let everyone know that the people around them are doing what they're supposed to be doing that they are converting. Here, they are also make it a bit of a competition by showing the "top" places. These nudges undoubtedly contributed to the 74% conversion rate.

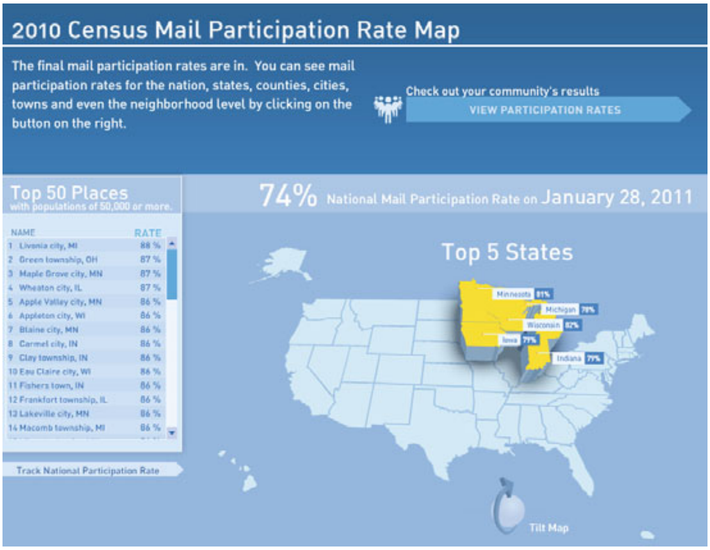


Fig. 2. U.S. Census Web Site – Response Rate by Region

3 Positive Persuasion

We must also dissuade people from the knee-jerk assumption that "persuading" is necessarily a bad thing, akin to "tricking." Is a hammer an evil tool? Is it a good tool? It is neither. It is simply a tool. It can be used for good or for evil. The principles of persuasion are tools. They can be used for good or for evil. You can persuade a child that eating fast food is cool and fun. Or you can persuade an elderly patient to adhere to their medication regimen. Both will use the principles of persuasion. One of them is using the principles for good and the other perhaps not.

For some, talking about "encouraging" is much more palatable – encouraging engagement, encouraging adherence, encouraging participation. No matter what we call it, though, we are trying to persuade someone to do something. For government or non-profit sites, we are typically not trying to persuade someone to buy something, we are trying to persuade them to take some other action. We are, however, trying to persuade them. We must be comfortable with the notion that persuasion is a perfectly acceptable method for achieving noble goals.

3.1 Obesity

Obesity is an epidemic plaguing America. It contributes to heart disease, diabetes, and a host of other secondary health issues. On top of the incredibly disturbing health implications for obesity, there are very significant costs to treat and respond to the health issues that it causes. We would be hard pressed to find a person that would argue that persuading people to lead a healthier lifestyle and avoid obesity. Let's look at one non-profit organization's (Shape Up America) web site [13] to see how they might be trying to persuade people to "shape up".

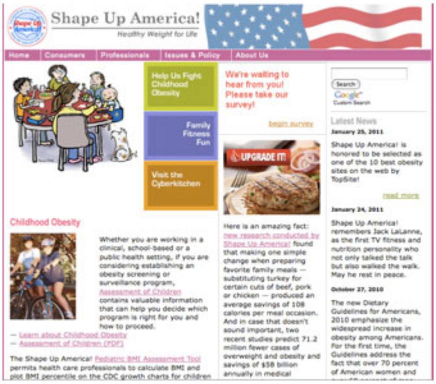


Figure 3a

Figure 3b

Fig. 3. Shape Up America Web Site

Figure 3a shows the top of the Home page. We see enticing imagery that speaks to our "lizard brain", showing us tasty food that we'd like to eat [3]. We also see imagery of "people like me" in the photos of the people riding the bicycles [4]. Figure 3b shows the bottom portion of the home page. We see other people sharing their successes and sharing their stories (social proof) [4]. Are any of these attempts to persuade people to avoid or manage obesity nefarious attempts to trick people? Of course not.

3.2 Quit Smoking

If obesity doesn't scare you enough, smoking should. Each year, more than five million people die because of smoking-related illnesses. SmokeFree.gov is a joint venture of the National Cancer Institute and several other government agencies. Smoking-related illnesses and deaths cost American tax payers almost two billion dollars per year [14]. It is the single most preventable cause of death for adults in America [15]. Figure 4 shows a page targeted at helping people who are thinking of quitting [16].

Does this design persuade you to quit, or to at least further explore the idea of quitting? I would argue no. There is a good amount of information on the page, but it doesn't highlight others who are being successful (i.e., "converting") like the MN tax or Census groups did. It isn't leveraging any social cues like the Shape Up America site does. There is plenty of opportunity to improve the persuasiveness of this design. Would increasing the persuasiveness be considered a bad thing? Would convincing more people to stop smoking be a good thing or a bad thing?

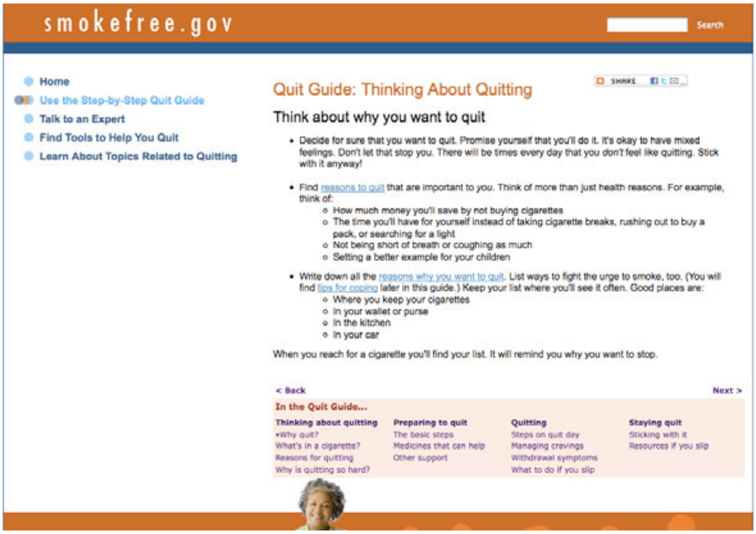


Fig. 4. Smokefree.gov Web Site

3.3 Awarding Savings

We've all been annoyed by signing up for services or getting spammed with emails or other things because we did not opt-out of something in an online form. When that happens, we feel like we're being tricked in to getting the service. Why should we have to do the extra work to NOT get something we don't want? One government employee won a trip to visit with the president for nudging users to accept an online version of a publication instead of a print one, thus saving the government a lot of money:

> *"Trudy Givens from Portage, Wisconsin works for the Bureau of Prisons. The Federal Register is currently mailed to her workplace and approximately 8,000 Federal employees every workday. Most of the interested public now accesses the Federal Register online. While statute requires that hard copies be available, allowing recipients to opt-in for hard copy delivery could yield savings associated with printing and postage. When a similar "opt-in" (with fee) option was offered to the public, the number of hard copies mailed was reduced from roughly 25,000 to 500 recipients."* [17]

So, this design change was done to persuade people to take a specific action. Was it done to trick anybody in to doing anything they didn't want to? No. It was done to save taxpayers many thousands of dollars while still allowing people to get a print copy if and when they want it.

4 Summary

Designing a successful web site means designing a successful user experience. Designing a successful user experience means creating a design that allows users to "can do" and "will do". Put another way, a successful user experience is one that is both usable and persuasive. Many designers of government web sites do not fully appreciate how persuasive design applies to their designs. Some think that they persuasion is only about persuading people to purchase things. Others think that "persuasion" is bad thing akin to trickery.

Every web site has a goal for its visitors – to buy, to engage, to act, to respond… to convert. Our goal as web site designers is to persuade people to convert to nudge them to action. We want to nudge them to pay their taxes, fill out their census surveys, lose weight, or quit smoking. Some of these things are things they might do while interacting with the web site just like they might do while interacting with an e-commerce web site. Many others, however, are things that they would do outside of the context of the site. If we broaden our understanding of what it means to convert and the role of persuasion, it is clear that persuasive design has as much of a role in the design of government and non-profit web sites as it does for retail.

References

1. Miller, K.: We Don't Make Widgets: Overcoming the Myths That Keep Government From Radically Improving. Governing Management Series (2010)
2. Thaler, R.H., Sunstein, C.R.: Nudge: Improving Decisions About Health, Wealth, and Happiness. Penguin Group (2008)
3. Weinshenck, S.: Neuro Web Design: What Makes Them Click. New Riders Press, Indianapolis (2009)
4. Human Factors International, HFI demonstrates the power of PET design, `http://www.humanfactors.com/home/PETdesignvideo.htm`
5. Cialdini, R.B.: Influence: The Psychology of Persuasion (Revised Edition). Collins Business Essentials (2007)
6. Norman, D.A.: Emotional Design: Why we love (or hate) everyday things. Basic Books, New York (2004)
7. The Minnesota income tax compliance experiment: State tax results, `http://mpra.ub.uni-muenchen.de/4827/`
8. Internal Revenue Service, `http://www.irs.gov/`
9. WI Department of Revenue, `http://www.revenue.wi.gov/`
10. TurboTax, `http://turbotax.intuit.com/`
11. H&R Block, `http://www.hrblock.com/`
12. 2010 Census Mail Participation Rate Map, `http://2010.census.gov/2010census/take10map/`
13. Shape Up America, `http://www.shapeup.org/`
14. Smoking deaths cost U.S. $92 billion per year, `http://usgovinfo.about.com/od/medicalnews/a/smokingcosts.htm`
15. Centers for Disease Control, Chronic Disease Prevention and Detection, `http://www.cdc.gov/chronicdisease/resources/publications/aag/osh.htm`
16. Smokefree.gov, `http://www.smokefree.gov/`
17. Nudge blog: Improving Health, Wealth, and Happiness, `http://nudges.org/`

An Experiment about How to Feel Temperature Change of Mouse

Shigeyoshi Iizuka

Kanagawa University
2946 Tsuchiya, Hiratsuka-Shi, Kanagawa, 259-1293 Japan
shigeiizuka@gmail.com

Abstract. Recently, some methods of using thermal information on computer interaction are considered. In order to utilize the thermal information more effectively, it is extremely important to understand how to feel the user to use the thermal information. A "mouse" is generally used when the computer is operated the index finger is used for it while coming in contact. Then, the thermal information is presented the point of the index finger through the mouse. The mouse is installed the Peltier device in the part that touches the tip of index finger, and can present the thermal information to the computer assisted user by warming and cooling the Peltier device. In the first experiment, it was experimented by ME (Magnitude Estimation) method to clarify user's how to feel it and the characteristic to the thermal stimulation. As a result, it was confirmed that the proportion of the degree of the change in user's sense to the degree of the change in presented thermal information. In the second experiment, it was experimented by using the control knob to understand the tendency of the time progress to how feel the user about the temperature. As a result, some features were found about how to the temporal variation of the temperature stimulus to feel the user.

1 Introduction

It has come to be able to send and receive various information by the development of the communications means using the computer. However, it is difficult to often communicate information only in the characters and the images, and to transmit slight feeling of quality of the object and human feeling efficiently. It is thought that the more two or more feelings are simultaneously used for the interface of a communication, it is friendly to users and an actual feeling increases. Therefore, it is thought that using various consciousnesses positively can support the communication like browsing and choice at the time of communicating information in daily life.

On the Web communication, users depend on the great portion of information acquisition on vision, so it is thought that being used in complement is good as for consciousness information other than vision. So, I make a proposal concerning the method of supporting communication like browsing and choice in communication using the computer by showing thermal information, as a way with which vision information is compensated.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 575–581, 2011.

In this research, it is focused on "mouse" used in many cases, using generally then the computer is operated as a device by which the user touches and contacts the palm, and the method of presenting thermal information as addition information on vision information through a mouse is adopted.

2 Thermal Information Presentation Method

2.1 Outline

The mouse function to perform scan of XY plane of the computer screen and two button inputs remains as it is, the mouse which can present the user while using computer the thermal information was made as an experiment (Fig. 1). The mouse is installed the Peltier device in the part that touches the tip of index finger, and can present the thermal information to the computer assisted user by warming and cooling the Peltier device, when the user clicks the contents (picture etc.) itself or a part of it which the thermal information was assigned on the computer screen. Therefore, the thermal information output function was added without the mouse spoiling the function of the mouse as an input device itself, this “mouse” is the general-purpose device which can be used only by USB connection, users can acquire additional information without the additional device.

Fig. 1. Appearance of Mouse

2.2 Architecture

As shown in Fig. 2, the mouse made as an experiment is made by being installed a Peltier device and a temperature sensor to a USB optical mouse, and makes it possible to output the thermal information to users with the Peltier drive and one tip micro-computer which were connected through the cable. The Peltier drive is embedded with the semiconductor temperature sensor near the right button of the USB optical mouse (for right hands). In the inside of the mouse, the optical mouse processing circuit is built with the heat dissipation board for Peltier. Since these are mounted using the space inside the mouse case, as shown in Fig. 1, appearance of the mouse is still the usual mouse except being installed with the Peltier drive, and the click of the right and left as a mouse, wheel operation, etc. can be used for or it like usual. The Peltier drive circuit and the control microcomputer are also connected with PC through the USB-serial conversion module. DC converter is built into this, and the electric power for Peltier device is acquired from USB electric power.

And, temperature control is executed according to the program written in the one tip microcomputer. Designated temperature is inputted from PC through USB serial, and the control direction (heating or cooling) is determined as compared with the information from the temperature sensor grounded at the Peltier device side.

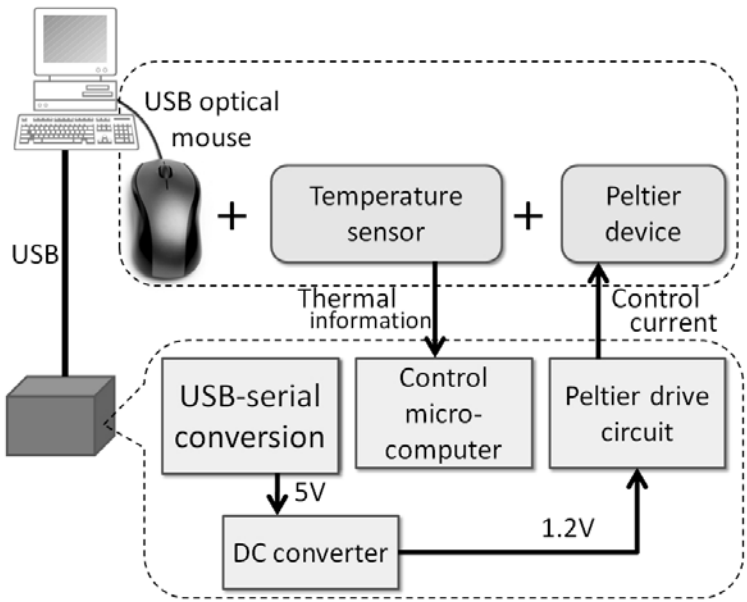


Fig. 2. Architecture

3 Experiment

In order to develop the technique of presenting thermal information as real service as a way with which vision information is compensated using the mouse stated for the preceding chapter, it is indispensable to understand the relation between presentation temperature and user's consciousness. Then, two experiments to clarify them were conducted. These two experiments are shown below.

3.1 Experiment

Then, it was assumed that user's how to feel it and the characteristics to the temperature stimulus were clarified by using “Magnitude Estimation method" that expressed a sense amount of psychology of the user to physical amount to stimulate and it numerically.

The conditions of the experiment were as follows.

- Participants : 10 men and women 20-30's
- Presentation stimulus: 8 stages (10.5 degrees C / 12.5 degrees C / 15 degrees C / 18 degrees C / 21.5 degrees C / 26 degrees C / 31.5 degrees C / 38 degrees C)
- The method of presenting stimulation: The mouse was gripped while unpresented, the experiment operator controlled the temperature of the mouse manually, and it set it to the stimulation (the above temperature) decided beforehand

3.2 Result of Experiment

First, regression line and formula which were drawn from the result of the experiment shown in Chapter 3 are shown in below.

Furthermore, the drawn power function is as follows.

$$y = 1.3102x + 0.1789$$
$$(R^2 = 0.9617)$$

The indicative power index number is as follows.

$$R = 1.509732 + S^{1.3102}$$

Then, ME values computed with the drawn power function are plotted on graph, and the characteristics is clarified. The logarithmic expression graph which both of the X-axis and the y-axis was displayed by logarithm is shown in Fig. 3. As a result,

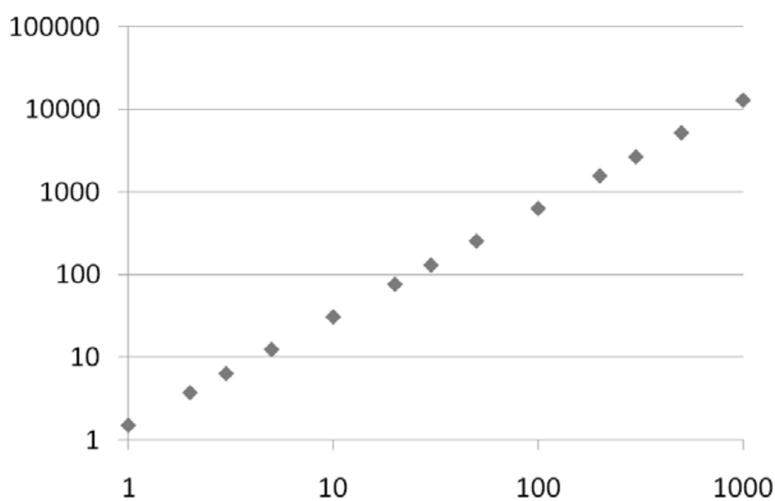


Fig. 3. Logarithmic expression graph

the temperature stimulus which user feels for an index finger with the mouse by presenting thermal information, it was confirmed that the proportion of the degree of the change in user's sense to the degree of the change in presented thermal information.

3.3 Experiment

In this experiment, it was experimented by using the control knob to understand the tendency of the time progress to how feel the user about the temperature. Concretely, it was carried out by how to operate a control knob with the left hand while gripping the mouse to the right hand. The participant expressed the state of the temperature felt in the right hand that touched the mouse by rotating the control knob. The details of the method of the experiment and the result are shown below.

Method

The conditions of the experiment were as follows.

- Participants : 10 men and women 20-30's
- Presentation stimulus: 7 stages (10 degrees C / 12 degrees C / 14.5 degrees C / 17.5 degrees C / 26.5 degrees C / 32 degrees C / 38.5 degrees C)
- Stimulus presentation procedure :
- Participant grasps the mouse in the state of temperature unpresenting
- (After mouse temperature is stabilized) The experiment management person controls the temperature of a mouse towards the stimulus (the above temperature) set up beforehand
- Maintain the temperature state near setting temperature for 20 seconds
- (After 20-second progress) Stop the temperature control of the mouse
- Then, maintain the state as it is for about 60 seconds

From a start to an end time of the above-mentioned presentation procedure, the participant expresses the current thermal feeling for the right hand with which the mouse is touched by rotating PowerMate(Fig. 4).

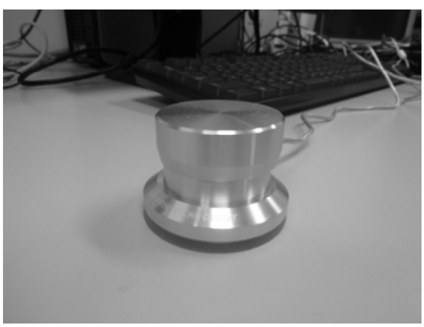


Fig. 4. PowerMate (Griffin Technology)

3.4 Result of Experiment

Some sample of graphs (only part) showing how each participant feels the thermal information of the mouse (numerical value) and the temperature of the mouse is shown in Fig. 5.

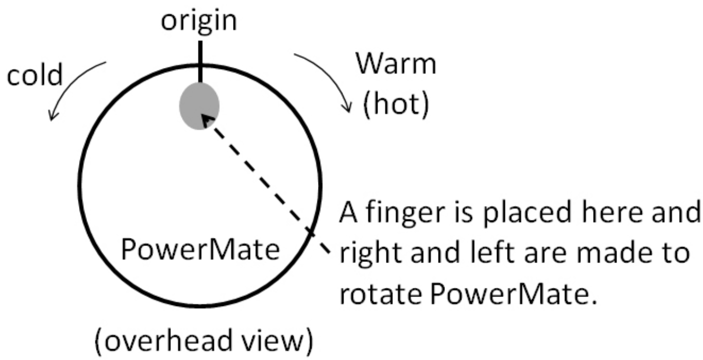


Fig. 5. Operation Method of PowerMate

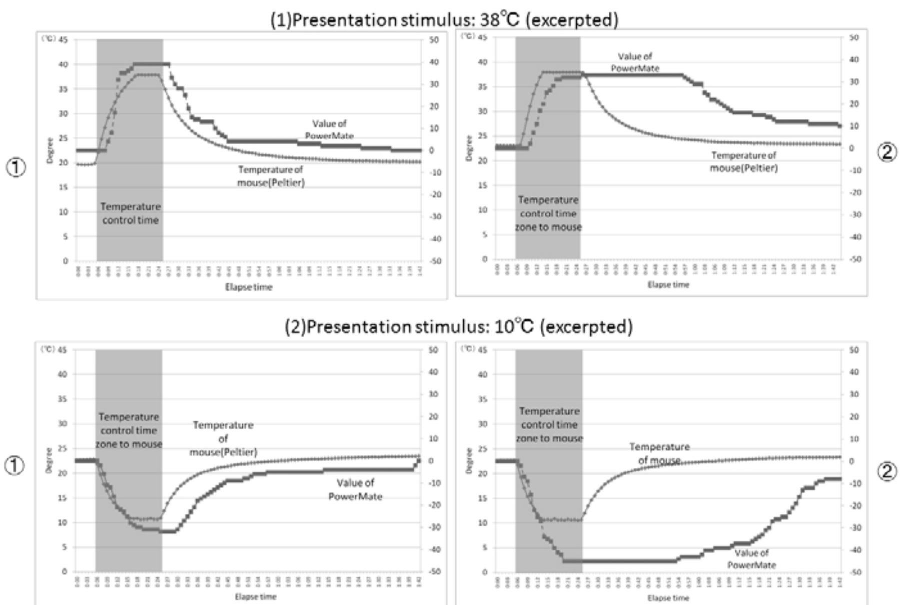


Fig. 6. Result of Experiment

As a result, some following features were found about how to the temporal variation of the temperature stimulus to feel the user:

- How to feel it changes by most synchronizing with the presented temperature change.
- How to feel it changes from the presented temperature change after some delay.

4 Conclusion

In this research, it is focused on "mouse" used in many cases, using generally then the computer is operated as a device by which the user touches and contacts the palm, and the method of presenting thermal information as addition information on vision information through a mouse is adopted. Actually a mouse was made as an experiment which is installed the Peltier device in the part that touches the tip of index finger, and can present the thermal information to the computer assisted user by warming and cooling the Peltier device.

In order to develop the technique of presenting thermal information, it is thought that it is indispensable to understand the relation between presentation temperature and user's consciousness. Then, two experiments to clarify them were conducted.

In the first experiment, it was experimented by ME (Magnitude Estimation) method to clarify user's how to feel it and the characteristic to the thermal stimulation. As a result, it was confirmed that the proportion of the degree of the change in user's sense to the degree of the change in presented thermal information. In the second experiment, it was experimented by using the control knob to understand the tendency of the time progress to how feel the user about the temperature. As a result, two following features were found about how to the temporal variation of the temperature stimulus to feel the user: “How to feel it changes by most synchronizing with the presented temperature change.” and “How to feel it changes from the presented temperature change after some delay”.

Hereafter, while acquiring knowledge about the optimal form of the device and the presentation method of thermal information, I aim to realize the technique of supporting communication by presentation thermal information as a way which vision information is compensated and to continue to give an actual feeling more to information on computer communication.

Clout: The Role of Content in Persuasive Experience

Colleen Jones

Content Science,
3455 Peachtree Road Atlanta GA 30326, USA
colleen@content-science.com

Abstract. For a variety of reasons, content has been excluded from the discussion of persuasive design, both in academia and practice. This paper argues that content is a missed opportunity to make a digital experience not only inform or instruct but also influence. I explain the causes and consequences of disregarding content, then define the proper role and benefits of content. To improve the results of a persuasive experience, content can and must have a central role in planning, executing, and evaluating the experience.

Keywords: persuasive design, influential content, content strategy, emotion, psychology, rhetoric, product strategy, marketing strategy, behavior change, attitude change.

1 Introduction

Results. From improving health, to selling products, to spreading good ideas, everyone wants results. But, many of us aren't getting them from the persuasive experiences we plan and create.

For more than 13 years, I've watched the interactive industry try the following same approaches to influence again and again, striving for results: [1]

- Using pushy tricks and optimizing small tweaks.
- Adding overpromised technology features.
- Focusing on attractive and usable design without substance or meaning.
- Using outdated marketing techniques, such as broadcasting a message.

If these approaches work so well, why are our results not better? For example, the global conversion rate index has hovered at 2-4% for the past 10 years. [1] Albert Einstein defined insanity as "Doing the same thing over and over and expecting different results." To get different results from our persuasive experiences, we need to try a different way. I'm convinced that way is content strategy, or planning for digital content.

2 A Brief History of Persuasive Design Ignoring Content

Content—the text, images, audio, and video that comprise an interactive experience—has been omitted from discussion in persuasive design academia and practice. While

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 582–587, 2011.

proving a negative is difficult, we can see some evidence of academics and practitioners viewing content as unimportant, beyond scope, or too difficult to tackle in research or practice.

2.1 In Persuasive Design Academia

Perhaps the most notable work in persuasive design lies with Stanford University psychologist B.J. Fogg. In his book *Persuasive Technology: Using Computers to Influence What We Think and Do*, Fogg implies the importance of content to persuasive games and persuasive social actors. [2] However, Fogg *explicitly* addresses design and technological capability throughout the book. In other words, having quality content seems largely assumed.

In a later academic article, Fogg recommends focusing on behavior, or action, over attitude as a key metric for change. [3] Because much content influences attitude more than action, this recommendation implies content is not essential to a persuasive experience.

Another example of academia omitting content from the persuasive design discussion is the article "Catalyzing a Perfect Storm" in *Interactions*. [4] While the article mentions several content-rich persuasive experiences, the article explains nothing about how content is or would be planned, maintained, and evaluated. Furthermore, its call for research funding includes areas such as usability but not content.

2.2 In Persuasive Design Practice

In design practice, content often is assumed to be someone else's responsibility—the client's responsibility or another business department's responsibility. As a result, content is left without ownership. Also, the terminology "design" tends to lead practitioners, clients, and stakeholders to think about visuals, not content. [5]

3 Consequences of Treating Digital Content as Filler

Treating content as the filler of a persuasive design usually means excluding content from persuasive design planning, creation, and evaluation. Excluding content this way causes many problems for everyone—from the designer to the user to the evaluator.

3.1 Problems for the Planner / Designer

Without considering content, a planner or designer for a persuasive experience will miss opportunities to influence the user (see 3.2). But, perhaps the most painful consequence of ignoring content is logistical. When a planner or designer assumes the client or another department will take care of content, the result often is project delay or failure. [5] Furthermore, if a process and resources are not in place to maintain and produce fresh content after the initial persuasive experience is launched, the experience will go stale quickly and risk losing its influential impact. [6]

3.2 Problems for the User

Content helps users gain a favorable attitude toward an action, motivate the user to an action, and guide users through or explain the action. When a persuasive experience focuses on design at the exclusion of content, the experience risks making a user feel pressured, manipulated, bored, or confused. For more about problems for the user experience, see *Clout: The Art and Science of Influential Web Content*. [1]

3.3 Problems for the Evaluator

An evaluator cannot assess what he or she cannot identify. Without a working vocabulary for content issues and techniques to analyze content performance, an evaluator cannot completely assess a persuasive experience. As Ginny Redish has noted, too many evaluations "focus only on finding information—not on how the information [content] itself works for people." [7] For a case study of evaluating *content* for Centers for Disease Control and Prevention, see "Testing Content: Early, Often, and Well." [8]

4 Industry Trends That Make Digital Content Critical *Now*

Recognized or not, content always has been important to an interactive experience. The timing of several trends in the interactive industry further underscores the importance of digital content. People are using online content to develop attitudes, make decisions, and take actions. These trends intensify the consequences of ignoring content and increase the urgency for addressing content in a persuasive experience.

4.1 Rise of Social Media

Social or new media includes blogs, user-generated content, and social networks such as Twitter and Facebook. Social media make publishing content easier—and more in demand—than ever before. At the same time, social networks are becoming a major gateway, in addition to search, into websites. [9]

4.2 Growth of Internet Use and Digital Business

The Pew Internet and American Life Project reports that all major market segments are using the Internet to decide everything from what to buy to what health treatment to take. [10] Forrester Research recently found that Americans, especially younger demographics, spend as much time using the Internet as they do watching television. [11] Online content, consequently, is a significant opportunity to influence people.

At the same time, more and more aspects of business—from marketing to purchase to support—are happening entirely online. For example, Mint.com helps people make financial decisions by bringing together a mix of personal data and a variety of content.

As well, business models such as media and retail are merging through content, as reported by Wall Street Journal. [12] For example, Groupon mixes a discount service with useful content about the products or services being discounted. As another example, retailers such as Pottery Barn, Williams-Sonoma, and REI are offering a range of media—from design tips to how-to videos—beyond typical product catalog content.

4.3 Rise of Mobile, Location-Based Services, and Ubiquitous Computing

Mobile offers new opportunities to repurpose content into a mobile website or application. Location-based services and ubiquitous computing provide new possibilities for content specific to a location, which is useful for travel destinations, college and university campuses, and large retail stores. For example, NC State University is piloting OnCampus, a location-based service that connects members of the university community with each other and with geolocation-based content. [1]

Together, these trends highlight the importance of content to persuasive interactive experiences. Avoiding content means missing the opportunities that these trends offer.

5 Academia and Industry Roots for Persuasive Content

While content has been left out of discussion in persuasive design circles, it has not been left out of the persuasion discussion altogether. We can turn to several academic disciplines and areas of business practice to ground content firmly in persuasion.

5.1 Academic Discipline and Business Practice

Persuasive content stems largely from the academic field of rhetoric. Rhetoric is the study of using language, and now other types of content, persuasively. Also pertinent to persuasive content is the field of cognitive psychology, or the study of how the brain processes information. Often, rhetoric and cognitive psychology are integrated into communication programs such as technical communication or writing.

The business practices related to persuasive content consist of marketing (including sales and advertising), public relations, and media / journalism. Partly because these functions occur more often online through content, content is becoming an increasingly influential force. (See 6.0)

For more about the academic and professional foundation of content strategy, including persuasive content, see *The Elements of Content Strategy*. [13]

6 Redefining Digital Content's Role in a Persuasive Experience

Academic and business roots help us envision more influential content. But, to make addressing content practical for persuasive experiences, we need to go further in redefining content's role.

6.1 From Filler to Core: Persuading from the Content Out

Rather than treat content as the "filler" of a persuasive design, it's time to treat content as the core. Content decisions affect the tone, flow, substance, meaning, and more of a persuasive experience. As Jeffrey Zeldman noted at An Event Apart in 2010, it's time to "design from the content out." This perspective has important implications for process and evaluation.

Process Implications. Content is critical to every step of planning and executing a persuasive experience. From developing the strategy to building and launching the experience, content must be a primary consideration, not an afterthought. And, after an interactive effort is launched, maintaining the content over time is critical. This new process calls for new artifacts such as content inventories, content matrices, content concepts, editorial style guides, editorial calendars, and more. For details about planning influential content, see *Clout: The Art and Science of Influential Web Content* and *The Elements of Content Strategy*. [1][13]

Evaluation Implications. If content is the heart of a persuasive experience, then no evaluation of such an experience is complete without examining the content. Evaluating content well in the planning and execution phases requires qualitative and quantitative research. And, after a persuasive experience is live, constant evaluation of the content's performance will provide insight into what content works well and what content needs adjustment. For a close look at evaluating the influence of digital content, see *Clout: The Art and Science of Influential Web Content*. [1]

6.2 Benefits

Restoring digital content to the core of persuasive experience offers three key benefits.

Better Practice. By addressing content in planning, creation, and evaluation, the practice of persuasive experience improves significantly. Projects are more likely to be completed at a high level of quality with a low level of risk. And, persuasive experiences with proper content planning will remain influential over time

Better Research. Restoring content to a central role opens a new door—or re-opens a previously closed door—to bodies of research about content and persuasive experience. Media studies, rhetoric studies, and more can inform this badly needed research into how digital content influences attitude and action.

Better Results. Improving the practice and research of persuasive experience will yield better results from them—whether those results are more sales or healthier people.

7 Conclusion

Although content has a firm academic and practical grounding in persuasion, it has been excluded from the discussion of persuasive design. Treating content as the filler of a persuasive design causes problems for the planners, users, and evaluators of a persuasive experience. When we overlook content this way, we miss an opportunity to make a digital experience influence the users' attitudes and actions. To improve results, content can and must have a central role in planning, executing, and evaluating a persuasive experience.

References

1. Jones, C.: Clout: The Art and Science of Influential Web Content. New Riders, Berkeley (2011)
2. Fogg, B.: Persuasive Technology: Using Computers to Change What People Think and Do. Morgan Kaufman, San Francisco (2002)
3. Fogg, B.J.: [book chapter]
4. Winchester III, W.: Catalyzing a Perfect Storm: Mobile Phone-Based HIV-Prevention Behavioral Interventions. Interactions XVI(6) (2009)
5. Halvorson, K.: Content Strategy for the Web. New Riders, Berkeley (2010)
6. MacIntyre, J.: The Day Two Problem. Atlanta Content Strategy (2010)
7. Redish, G.: Letting Go of the Words blog, `http://www.redish.net/writingfortheweb/`
8. Jones, C.: Testing Content: Early, Often, and Well. IA Summit (2010)
9. Gibs, J.: Social Media: The Next Great Gateway for Content Discovery. Nielsen Net Ratings (2010)
10. Change in Internet Access by Age Group 2004-2009. Pew Internet and American Life Project (2010)
11. Rousseau-Anderson, J.: US Consumers Now Report Spending Equal Time with TV and the Internet. Forrester Research (2010)
12. Fowler, G., Vascellaro, J.: Media and Retail Firms Meld Business Models Online. Wall Street Journal (2010)
13. Kissane, E.: The Elements of Content Strategy. A Book Apart, New York (2011)

Influencing Mechanism of Apparent Space Dimensions on Interface Aesthetics and Apparent Usability

Tian Lei[1], Yingbin Zhou[2], Xiang Li[3], and Xiaoli Chen[4]

[1] Dept. of Industrial Design, Huazhong University of Science and Technology, 430074 Wuhan, P.R. China
[2] School of Art and Design, Wuhan Institute of Technolgy, 430073 Wuhan, P.R. China
[3] School of Art and Design, Wuhan University of Technology 430070 Wuhan, P.R. China
[4] Faculty of Mechanical and Electronic Information, China University of Geosciences 430074 Wuhan, P.R. China
andrew.tianlei@gmail.com

Abstract. Apparent usability (AU) and interface aesthetics are the two important factors in HCI, which are affected by the apparent space dimension (ASD). This paper, by making two experiments, explored the influencing mechanism of ASD on them. The results show that: 1) AU is made up of subjective feelings, operation, and cognition; 2) interface aesthetics is made up of impression beauty, material beauty, and hominine beauty; 3) participants' subjective feelings increase with the addition of ASD; 4) participants have the strongest operational ability in the apparent 2-dimensional space but the weaker in the other two; 5) participants' cognition for the interactive system decrease with the addition of ASD; 6) interface's impression beauty increases with the addition of ASD; 7) interface's material beauty and hominine beauty are both the best in the apparent 2-dimensional space, but not good enough in the other two.

Keywords: apparent space dimension, apparent usability, interface aesthetics.

1 Introduction

Apparent Usability (AU) is an ignored branch of the Usability, which are the users' feelings or sensations about whether products or an interactive system is easy to use before using it. Just like Usability, AU has effect on users' acceptance and evaluation of an interactive system.

Aesthetics is another important factor in HCI. Recent work has suggested that aesthetics plays a considerable role in the interactive system. When a user faces an interactive system, the formal elements and their semantics will stimulate him immediately and cause him to produce perceptions and emotions, such as the overall impression, degree of preference, and interaction attitude[1],[2]. Gait reported that more interesting interface could increase a user's arousal and sustain his effectiveness and interest in the interactive system [3]; Jordan, with his experiments, found that aesthetics was the important determinant of whether an experience was pleasurable or not. Norman, D.A. thought that aesthetic design could affect a user's preference more than the traditional design based on Usability and operation performance [4].

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 588–597, 2011.

Aesthetics can improve the interactive system's AU. For instance, Kurosu and Kashimura took the ATM's interface layout as independent variable, and apparent usability, inherent usability, actual usability, and aesthetics as dependent variables, and then made an experiment. The result indicated that there was a considerably high correlation between apparent usability and visual image (interface aesthetics) [5], [6].Tractinsky improved this experiment and got the similar results [7]. However, some other research finds that the relationship between them is not so. For example, Angeli etc. [8]found that the metaphor-based interface was better than the menu-based one in the expressive aesthetic dimension and users preferred the interface having a great attraction in the expressive aesthetic dimension although its inherent usability and apparent usability were less than the menu-based interface's.

Is there a potential factor that could affect AU and interface aesthetics at the same time and make the positive or negative correlation between them under different conditions? Through the initial screening, the factor of apparent space dimension (ASD) was brought into our study. ASD is the understanding of the number of interface's space dimensions, which is caused by the four key design techniques, gradients, shadows, perspective and directions of the movement. This paper, by making two experiments, will research the structure of AU, the influence of ASD on AU, and the relationship among ASD, AU, and interface aesthetics.

2 Experiment

This experiment is to research AU's structure and the influence of ASD on AU.

2.1 Variables

According to the four critical design techniques, the independent variable, ASD, was divided into three levels and each level had four classical values (layout), as shown in Table 1 and Fig. 1.

Table 1. Parameters of ASD

Level	Gradients	Shadows	Perspective	Directions of the movement
Level 1	×	×	×	X or Y
Level 2	√	×	×	X and Y
Level 3	√	√	√	X, Y and Z

The dependent variable was AU. It was composed of several different facets and each was corresponding to a question. According to Hassenzahl's 7-facet questionnaire [9] and Peiwen Lin's 6-facet questionnaire [10], three facets of pleasure, innovation and interest were involved into this experiment. The integral questionnaire is seen in Table 2.

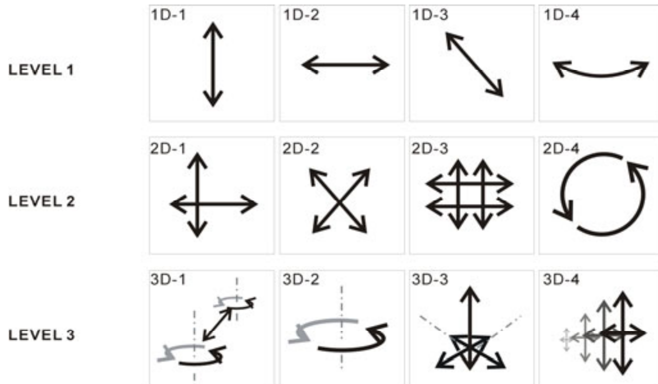


Fig. 1. Each value of the independent variable

There are many factors affecting the dependent variable. Here, icons' shapes, colors, texture, semantics, pix, and slide speed were the control variables which were controlled.

Table 2. The questionnaire of AU

No.	Questions	No.	Questions
Q1	Is it comprehensible?	Q7	Is it familiar?
Q2	Is it simple?	Q8	Is it interesting?
Q3	Is it ease- to-use?	Q9	Does it give you the passion of usage?
Q4	Is it clear?	Q10	Is it creative?
Q5	Is it safe?	Q11	Is it impressive after using it?
Q6	Is it controllable?	Q12	Does it give you the pleasant user experience ?

2.2 Materials and Participants

Based on the above variables, 12 experimental samples were designed, as Fig. 2 shows. Each sample was corresponding to a value. Fifty-six college students (20-25 age), majoring in design, took part in this experiment.

3 Results and Discussions of the Experiment

3.1 Structure of AU

A PCA was conducted for the data from AU experiment (Table 3). Three main factors were extracted which could explain 38.996%, 28.209%, and 24.547% of the total variance respectively after rotation. It suggests that participants' differences in AU caused by ASD can be totally explained 91.753% by these three main factors. The loadings of Q7, Q8, Q10, Q11, and Q12 on factor 1 were outstanding, which were 0.822, 0.929, 0.9, 0.934 and 0.819 respectively. These facets mainly reflect the participants' subjective understandings of interactive systems. So, it can be named subjective feelings factor (SFF). The loadings of Q1, Q3, Q6, and Q9 on factor 2 were bigger, which were 0.752, 0.842, 0.846 and 0.853 respectively. These questions are closely correlated with users' operation behaviors, and accordingly factor 2 can be named operation factor (OF). The contributions of Q2, Q4, and Q5 to factor 3 were bigger, and their factor loadings were 0.917, 0.834 and 0.927 in turn. These AU facets are closely related to users' cognitive loadings and information amount. So it can be called cognition factor (CF). Thus, we find that AU is mainly made up of three mutually independent components which are the subjective feelings factor, the operation factor and the cognition factor.

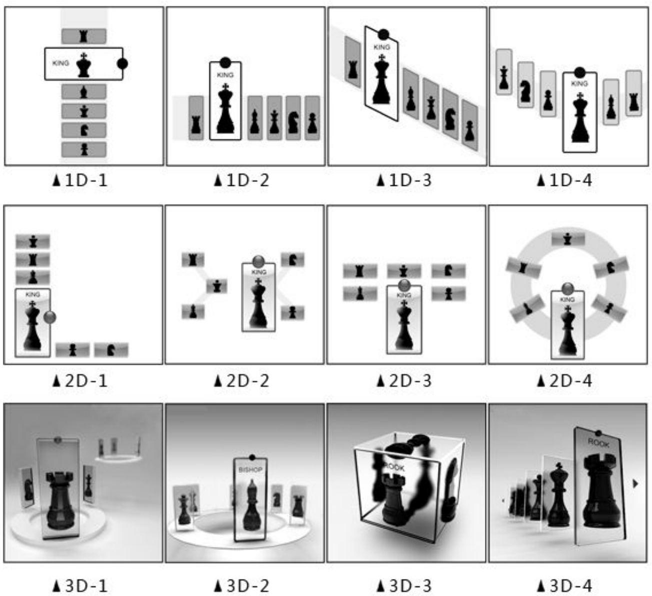


Fig. 2. Experimental samples

Table 3. Results of the Principal Component Analysis of AU facets

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.922	57.682	57.682	4.680	38.996	38.996
2	3.092	25.763	83.445	3.385	28.209	67.206
3	.997	8.309	91.753	2.946	24.547	91.753

3.2 Effect of ASD on AU

What on earth does ASD influence AU? Taking the regressors of the AU's facets on three main factors as the new variables, seen in Table 4, and alternately choosing two factors from the three as coordinate axes, we got the samples' spatial distributions of AU, as shown in Fig.3, and Fig. 4.

Table 4. Regressors of the AU's facets on the three main factors

Independent variable	SFF	OF	CF
1D-1	-.31457	-1.05665	1.09496
1D-2	.28325	-.25887	1.67797
1D-3	-1.62896	-.68036	.38898
1D-4	-.99274	-1.12577	.15391
2D-1	-.38316	.63588	-.44406
2D-2	-1.38397	.27236	-2.04399
2D-3	.19673	1.28346	.89421
2D-4	-.11937	1.97979	.17212
3D-1	1.23741	-.66683	-.94712
3D-2	.61676	.88486	.06972
3D-3	1.41528	-.70276	-.81447
3D-4	1.07333	-.56511	-.20222

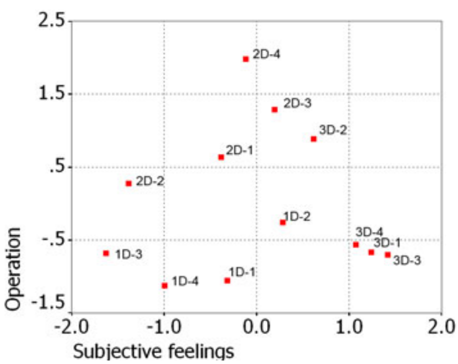


Fig. 3. The samples' distributions in SFF-OF space

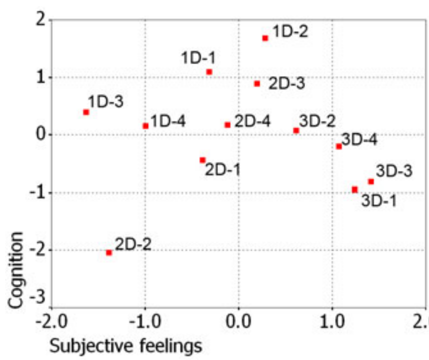


Fig. 4. The samples' distributions in SFF-CF space

Fig.3 shows that: 1) Only the 2D-3 and 3D-2 two samples have the positive values on both SFF axis and OF axis, and only 1D-3, 1D-4, and 1D-1 are negative on the two axes; 2) the samples of 1D-2 3D-1 3D-3 3D-4 are all positive on SFF axis, and 3D-1, 3D-3, and 3D-4 are the top three on this axis, but their values on OF axis are all less than zero; 3) the samples of 2D-1 2D-2 and 2D-4 are located in the positive part of OF axis, and the value of sample 2D-4 is the highest in all the samples on this axis.Yet, they have a weaker performance on SFF axis.

Fig.4 shows that: 1) 1D-2 2D-3 and 3D-2 are all located in the positive part on both SFF axis and CF axis, and the cognition of sample 1D-2 is the best in all the samples. However, either the sample 2D-2 or the one 2D-1 has a negative value on the two axes, and the 2D-2 has the worst cognition; 2) 1D-1, 1D-3, 1D-4 and 2D-4 have a good performance on CF axis but a weak one on SFF axis; 3) The values of 3D-2, 3D-3 and 3D-4 are all less than zero on CF axis, but greater than zero on SFF axis.

Averaging each sample's new variables according to the number of ASD, we got Fig.5, Fig.6 and Fig.7. From it, we find that:

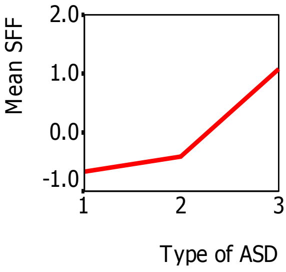


Fig. 5. The effect of ASD on SFF

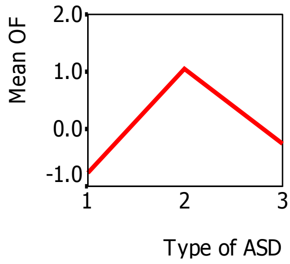


Fig. 6. The effect of ASD on OF

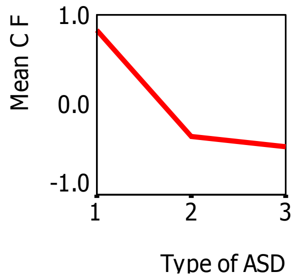


Fig. 7. The effect of ASD on CF

1) The participants' subjective feelings caused by interface increase with the addition of the number of ASD.

2) The participants have the strongest confidence in controlling their operations in the apparent 2-dimensional space but the weaker in the apparent 1-dimensional space and the apparent 3-dimensional space.

3) The participants' understandings of the interactive system at the cognitive level decrease with the addition of the number of ASD

4 Experiment

The purpose of Experiment is: 1) to verify the interface aesthetics' structure, and 2) to study the relationship between ASD and interface aesthetics. In order to find the influencing mechanism of ASD on both AU and interface aesthetics under the same conditions, we set up independent variables, control variables, experimental materials, and participants the same as Experiment 's. Here, we used semantic differential to capture interface aesthetics.

30 descriptive aesthetic vocabularies (DAVs) were the dependent variables in this experiment, as shown in Table 5.

Table 5. The dependent variables of experiment

No.	variable	No.	variable	No.	variable	No.	variable
1	Harmonious	9	Faddish	17	Steady	25	Mysterious
2	Unique	10	Exalted	18	Dreamlike	26	Classical
3	Natural	11	Smooth	19	Refreshing	27	Decorative
4	Elegant	12	Exquisite	20	Fine	28	Plain
5	Avant-garde	13	Feminine	21	Masculine	29	Traditional
6	Fresh and clean	14	Magnificent	22	Stately	30	Massive
7	Concise	15	Modern	23	Delicate and elegant	/	/
8	Warm	16	Rough	24	Lovely	/	/

5 Results and Discussions of the Experiment

5.1 Structure of Interface Aesthetics

The previous study has shown that interface aesthetics is made up of impression beauty, material beauty and hominine beauty. It is confirmed by this experiment again. Data from Experiment was dealt with PCA. Three main factors were extracted, and after rotation they respectively explained 46.005%, 19.78% and 18.07 % of the total variance produced by ASD, seen in table 6. Factor 1 was positively influenced by variables of "Unique", "Elegant", "Concise", "Faddish", "Exalted", "Smooth", "Feminine", "Magnificent", "Modern", "Dreamlike", "Fine", "Lovely", "Mysterious", "Classical" and "Decorative" because the factor loadings of these DAVs on factor 1 were all greater than 0.7. They were 0.953, 0.789, 0.889, 0.982, 0.952, 0.814, 0.906, 0.866, 0.952,0.862, 0.871, 0.914,0.873, 0.853, and 0.931 respectively. They convey a kind of formal beauty produced by association and empathizes. So, factor 1 is named impression beauty (IB). Factor 2 was affected by DAVs of "Harmonious", "Natural", "Fresh and clean", "Warm", "Exquisite", "Steady", and "Refreshing", whose factor loadings were 0.774, 0.909, 0.914, 0.811, 0.813, 0.685, and 0.611 respectively. These DAVs show a beauty caused by the physical situation and texture. So, factor 2 is called material beauty (MB). Factor 3 was influenced by "Avant-garde", "Rough", "Masculine", "Stately", "Plain", and "Traditional", and is named hominine beauty (HB).

Table 6. Results of the PCA of experiment

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	16.220	54.067	54.067	13.802	46.005	46.005
2	6.903	23.010	77.076	5.934	19.780	65.785
3	2.959	9.864	86.941	5.421	18.070	83.855

5.2 Effect of ASD on Interface Aesthetics

What on earth does ASD affect interface aesthetics? Taking the regressors of the DAVs on the three main factors as new variables, seen in table 7, and alternately selecting two from the three as coordinate axes, we got the samples' spatial distributions of interface aesthetics, seen in Fig. 8, and Fig. 9.

Then averaging each sample's new variables according to the number of ASD, we find the clear relationship between ASD and interface aesthetics (Fig. 10, Fig. 11 and Fig. 12). As follows:

1) The impression beauty of interface increases as the number of ASD rises

2) The material beauty and the hominine one are both the best in the apparent 2-dimensional space, and not good enough in neither the apparent 1-dimensional space nor the apparent 3-dimensional space.

Table 7. Regressors of the DAVs on the three main factors

Independent variables	IB	MB	HB
1D-1	-.77819	-1.37461	-.54331
1D-2	-.54677	.69628	-1.78656
1D-3	-1.00528	-1.33860	.32201
1D-4	-1.16931	.24226	-1.24642
2D-1	-.93940	.52454	1.06985
2D-2	-.98243	.05102	1.46975
2D-3	.27215	1.37063	1.06031
2D-4	.41976	1.54666	.18669
3D-1	1.12506	.14280	-.75302
3D-2	1.08128	.20675	-.65995
3D-3	1.28975	-1.14721	.41238
3D-4	1.23340	-.92052	.46827

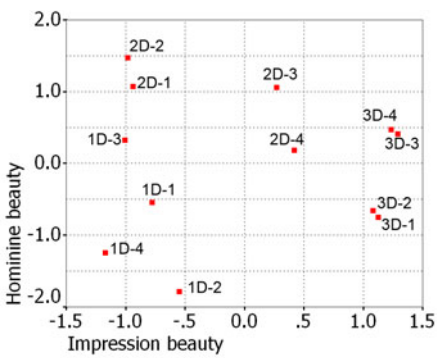


Fig. 8. The samples' distribution in IB-HB space

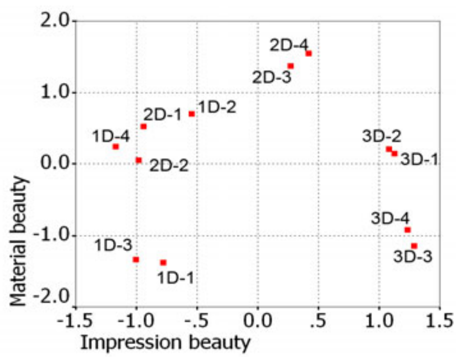


Fig. 9. The samples' distribution in IB-MB space

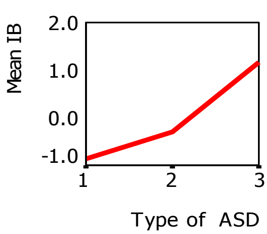


Fig. 10. The effect of ASD on IB

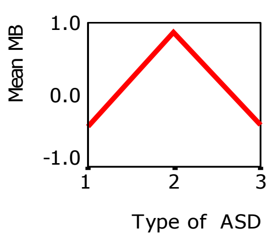


Fig.11. The effect of ASD on MB

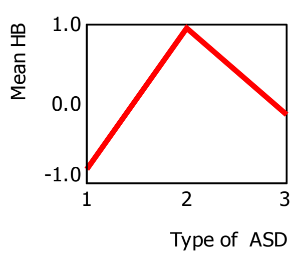


Fig. 12. The effect of ASD on HB

6 Summary

The above two experiments manifest that if changing the objects' gradients, shadows, directions of the movement, and perspective on interface, that is, increasing or decreasing the number of ASD, an interactive system's AU and aesthetics can make a different change accordingly. It seems that ASD determines the tones of AU and interface aesthetics.

How does interface aesthetics have effect on AU if the interactive system's ASD is decided? A Correlation Analysis was conducted among the six new variables, seen in Table 8. From it, we find that only between some specific factors does a statistically significant correlation occur. Impression beauty and subjective feelings are significant positive correlation at the 0.01 level (2-tailed), whose correlation coefficient r is 0.893. Material beauty and operation factor of AU are also significant positive correlation at the 0.01 level (2-tailed), whose correlation coefficient r is 0.772.

Table 8. Results of the Correlation Analysis

		IB	MB	HB	SFF	OF	CF
IB	Pearson Correlation	1	.000	.000	.893(**)	.171	-.172
	Sig. (2-tailed)	.	1.000	1.000	.000	.595	.593
MB	Pearson Correlation	.000	1	.000	.007	.772(**)	.121
	Sig. (2-tailed)	1.000	.	1.000	.983	.003	.707
HB	Pearson Correlation	.000	.000	1	-.194	.400	-.538
	Sig. (2-tailed)	1.000	1.000	.	.546	.198	.071
SFF	Pearson Correlation	.893(**)	.007	-.194	1	.000	.000
	Sig. (2-tailed)	.000	.983	.546	.	1.000	1.000
OF	Pearson Correlation	.171	.772(**)	.400	.000	1	.000
	Sig. (2-tailed)	.595	.003	.198	1.000	.	1.000
CF	Pearson Correlation	-.172	.121	-.538	.000	.000	1
	Sig. (2-tailed)	.593	.707	.071	1.000	1.000	.

These determined correlations provide a theoretical basis to enhance the usability of an interactive system quickly and effectively. For example, by adjusting interface's texture and impression beauty, a system's usability can be improved.

References

1. Schenkman, B.N., Jonsson, F.U.: Aesthetics and Preferences of Web Pages. J. Behavior and Information Technology 19(5), 367–377 (2000)
2. Heijden, H.V.D.: Factors Influencing the Usage of Websites: the Case of A Generic Portal in the Netherlands. J. Information and Management 40, 541–549 (2003)
3. Gait, J.: An Aspect of Aesthetics in Human–Computer Communications: Pretty Windows. J. IEEE Transactions on Software Engineering 11(8), 714–717 (1985)
4. Norman, D.A.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books, New York((2004)

5. Kurosu, M., Kashimura, K.: Apparent Usability vs. Inherent Usability: Experimental Analysis on the Determinants of the Apparent Usability. In: Conference on Human Factors and Computing Systems, pp. 292–293. ACM Press, New York (1995)
6. Kurosu, M., Kashimura, K.: Determinants of the Apparent Usability. In: Proceedings of IEEE SMC, Vancouver, Canada, pp. 1509–1514 (1995)
7. Tractinsky, N.: Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In: ACM CHI Conference Proceedings on Human Factors in Computing Systems, pp. 115–122. ACM Press, New York (1997)
8. Angeli, A.D., Sutcliffe, A., Hartmann, J.: Interaction, Usability and Aesthetics: What Influences Users' Preferences? In: 6th ACM Conference on Designing Interactive Systems, pp. 271–280. ACM Press, New York (2006)
9. Hassenzahl, M., Platz, A., Burmester, M., Lehner, K.: Hedonic and Ergonomic Quality Aspects Determine Software's Appeal. In: CHI 2000 Conference Proceedings, pp. 201–208. ACM Press, New York (2000)
10. Lin, P.W.: The Study of E-commerce Website Interface Design–Aesthetics vs. Usability, http://nctur.lib.nctu.edu.tw/handle/987654321/9860

The Health Machine: Mobile UX Design That Combines Information Design with Persuasion Design

Aaron Marcus

Aaron Marcus and Associates, Inc., 1196 Euclid Avenue, Suite 1F,
Berkeley, CA, 94708 USA
Aaron.Marcus@AMandA.com

Abstract. The author's firm combined information design with persuasion design to design a mobile phone application intended to change people's behavior about diet and exercise. The objectives were to change people's behavior and to avoid obesity and diabetes. The paper describes the user-centered user-experience development.

Keywords: culture, design, development, diet, exercise, health, information, nutrition, persuasion, social networks, user interface, user experience.

1 Introduction

Obesity and resultant Type 2 diabetes are major health concerns in the United States of America (US) and throughout the world. For those people wishing to maintain or improve their health, currently existing health-oriented mobile-phone applications provide many usable, useful functions, including medicine-intake monitoring, pain management, weight loss programs, exercises tutoring, and health-indicators tracking (*e.g.*, blood pressure and heart rate). According to the American Diabetes Association, a combination of exercises, healthier food, and weight control will be more effective to alleviate or prevent diabetes. Unfortunately, seldom do the applications combine many of these functions; they tend to be specialized. Above all, these products do not provide an overall "persuasion path" to change users' short-term and long-term behavior, leading to a healthier lifestyle.

2 The Health Machine Concept

Aaron Marcus and Associates, Inc. (AM+A) embarked on the conceptual design of a mobile-phone-based product, the Health Machine, intended to address this situation. The Health Machine's objective is to combine information design and visualization with persuasion design to help users achieve their health objectives, especially regarding obesity and diabetes by persuading them to adapt their lifestyle to include healthy eating and appropriate exercise. The Health Machine is intended especially for those people who have low income and less education, and thus may have less awareness, knowledge, and access to information about obesity and diabetes. AM+A intends,

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 598–607, 2011.

through the useful, usable, and appealing mobile device application, that the targeted users can be motivated to successful regimens of weight control and exercise, and to learn about and maintain a healthier life-style in the long-term. The Health Machine aims to answer these two critical questions: How can information design/visualization present persuasive information to promote sustainable, short-term and long-term health-behavior change? How can mobile technology assist in presenting persuasive information and promote behavior change of low income and less educated people?

3 Personas

Personas/user profiles are characterizations of primary user types intended to capture essential details of their demographics, contexts of use, behaviors, motivations, and their impact on the design solution. For the Health Machine, AM+A defined these:

Persona 1: Alan Marx, 67, Owner, Small Business. Mr. Marx was diagnosed with Type 2 Diabetes at the age of 66 after being borderline diabetic for several years. The condition is a mixture of heredity, lack of sufficient exercise, and eating portions that are too large. His weight climb until in October 2009, it reached 238 lb. for someone 6’1” in height. After being diagnosed with diabetes, assigned the use of a glucose meter, “required” to attend five classes on diabetes and nutrition, and urged to lose weight and increase exercise, he embarked on a sudden, massive change in health maintenance. He was motivated by fear of death and desire to achieve a workable goal. He succeeded in losing 50 pounds (down to 188, three pounds from his target goal of 185) in three months by reducing his calorie intake to about 1500 calories, noting on paper his daily nutrients, and increasing his daily exercise: to one hour on a treadmill at 4 mph, 100 sit-ups, and variable weight-lifting (15-50 lb.) three times per week. Since his initial achievement, he has become tired of constant data entry, and his weight has drifted upwards. He seeks to keep his weight below 200 lb. He wants a mobile device that will be easier for him to monitor his nutrients, log his exercise and glucose readings, and visualize his data, which was not possible with his manually-entered data.

Persona 2: Anna White, 57, Head of single-parent, multi-generational family Anna White is an African-American grandmother with a grade-school education who has worked as a domestic most of her life. She now weighs about 200 lb and is five feet two inches. She heads a single-parent family taking care of two children and three grandchildren. She has been diagnosed with Type 2 diabetes but is not very well informed about her condition, its causes, and options available to her through a community medical care-center.

Persona 3: Manuel Jiminez, 70, Husband, Founder of a gardening and lawn-care service. Manuel Jiminez is a legal immigrant from Mexico who started a gardening and lawn-care service now mostly managed and run by his son. He suffers from Type 2 diabetes, weighs approximately 250 lb., and is five-feet, six inches tall. He finds he doesn’t have time to do the exercise he needs to add to his daily schedule because of continuing work pressures to contribute to the family income.

4 Use Scenarios

Based on the user profiles cited and an interview with one actual person similar to Persona 1, AM+A determined use scenarios comprising the following tasks: Enter nutrient data with least effort, e.g., scanning food data labels, scanning restaurant offerings data labels, and going online to collect appropriate nutrient data.

- Review current and past data in table and chart modes, especially for new trends or for information that might change filters or goal targets for nutrition or exercise.
- Share data with other family, friends, physicians, groups, and "competitors."
- Read/react to communications about one's data, status, or trends.
- Compare one's own data with friends, with "stars" or heroes/heroines.
- Read/see information about proper food preparation, how to balance low fat and low salt content with good taste.
- Read/see exercise information, e.g., body positions for particular muscle groups.
- View future implications of current behavior.
- Read/see examples of equipment or food or restaurants that might be of interest.
- Upload/download photos relevant to progress as emotion/documentation markers.

5 Research

Comparison Study. Before undertaking conceptual and visual designs, AM+A first studied approximately 20 highly reviewed health iPhone applications. Screen comparison and customer review analysis, contributed to improvements of initial ideas for the Health Machine's detailed functions, data, information architecture (metaphors, mental model, and navigation) and look-and-feel (appearance and interaction). These iPhone applications [Marcus, 2011] are the following:

Diabetes-Related	**Nutrition**	iMuscle
Diabetes Log	Nutrition Menu	C25k Couch to 5K
Diabetes Planner and Carb Counter	Lose it!	
		Other Health
Diabetes Pilot	**Exercise**	BMP Calculator
Glucose Buddy	iFitness	Water your Body
Diabetes Helper 3.2	LiveStrong.com	Sleep Machine

Research Results. AM+A concluded that usable, useful, and appealing user-interface (UI) design must include incentives to lead to behavior change. Good, health-oriented mobile-phone applications should provide group comparisons, charts, illustrations, goals, competition, and/or step-by-step instructions to motivate people to change their behavior. The Health Machine needs to combine persuasion theory, provide better incentives, and motivate users' to achieve short-term and long-term behavior. Large, up-to-date databases are required. Users always demand large and searchable database, especially databases that are customizable. Customizable, flexible databases, by which users and their network of family and friends can easily add more information, are critical as a factor in increasing usage and an inevitable competitive advantage. In addition, in comparison to traditional manual data-input methods, a multiple data-entry system including label scanning and database searching is required to facilitate users' data-input process. Also, the device must encourage and strengthen team-oriented behavior change. Based on studies of persuasion theory, we discovered that team-oriented social comparison is a superior incentive for behavior change. Co-operation and competition within and among teams can encourage people to exercise more and carry out better diet control. Virtual rewards (*e.g.*, "stars", new skins for blogs, *etc.*) and real financial rewards (one corporate study [Associated Press, 2010] showed $500 can make a significant difference) can both be provided as motivation incentives for teams.

Last but not least: the Health Machine should be fun to use. Well-designed games will serve as an additional appealing incentive to teach, to train, and to inform users about how to select meal combinations wisely, how to exercise efficiently and effectively, and other techniques of nutrition and exercise. Also, the Health Machine should allow users to share their experience with friends, family members and the world, primarily through Facebook, Twitter and blogs.

6 Persuasion Theory

Based on Fogg's persuasion theory [Fogg, B. J., and Eckles, D., 2007] to create behavioral change, we defined five key attributes for the application:

Increase frequency of using application

- Motivate changing some living habits: work out, food choices, weight control
- Teach how to change living habits
- Persuade users to change living habits (short-term change)
- Persuade users to change life-style (long-term change)

Motivation is a need, want, interest, or desire that propels someone in a certain direction. From the sociobiological perspective, people in general tend to maximize reproductive success and ensure the future of descendants. We apply this theory in the Health Machine by making people understand that every action has consequences on their health condition change and their future.

We also adapted Maslow's *A Theory of Human Motivation* [Maslow, 1943], which he based on his analysis of fundamental human needs:

- Safety/security need is met by visualizing the amount of food expense saved.
- Belonging/love need is expressed through friends/family and social sharing/support.
- Esteem need is satisfied by social comparisons that display weight control and exercise improvements, and by self-challenges that display goal accomplishment.
- Self-actualization need is fulfilled by being able to visualize progress of health indexes and moods, and by predicting the change of the users' future health.

6.1 Impact on Information Architecture

Increase Use Frequency. Games and rewards are the most common methods to increase use frequency. AM+A developed a prototype pet training game and a meal combo game. In terms of the rewards, users will be awarded by both virtual rewards (such as “star” nominations and new skins for blogs) and real money rewards. In addition, we chose social comparison as another incentive to increase use frequency. Users may form groups with families/friends and participate in competitions.

Increase Motivation. The users’ future health condition prediction is understandably an important. By viewing their current health conditions and predicted future scenarios in the next 20-30 years, users will have a stronger impression and awareness. Because setting goals helps people to learn better and improves the relevance of feedback, the Health Machine asks users how much expense budget they want to save on healthier food, what calories/blood glucose level they want to achieve, and how much work-out they want to accomplish. Users receive suggested action plans to achieve each goal. We also created 10 monthly challenges. The incentive of self-accomplishment will encourage users to achieve step-by-step changes that will generate long-term changes. Social interaction also has an important impact on behavior change. The Health machine leverages social networking and integrated features like those found in forums, Facebook, Twitter or blogs. Users can send notes or messages to their social groups and share ideas with other people. The social ties will serve as an additional incentive to motivate behavior change.

Improve Learning. To improve learning, the Health Machine integrates contextual tips to explain how to eat healthier and increase exercise, the complications associated with diabetes, and ways to cope with too hi/low glucose levels. Users can also update latest research articles and news about diabetes and obesity. We also seek to make the education process both informative and entertaining. Games are proposed to teach users how to choose the right proportion, amount, and type of food for each meal without boring them.

7 Information Architecture

The Health Machine’s theoretical information architecture components, which appear in the accompanying figure are explained below.

My Condition. Provides timely information about food in order to enable the user to make healthy choices. The application will also assist in the recording of food nutrition consumption, such as calories, glucose, carbohydrates, *etc*., and making a record that can be of use to the patient as well as physicians, nurses, dieticians, and other healthcare service providers. Exercise is another crucial factor in the equation for controlling weight. To help patients with obesity and diabetes achieve and maintain a healthy body condition, the Health Machine supports the development of healthy exercise habits, training and teaching users with the appropriate means and amount, and enabling the integration of new activities into the user’s lifestyle in a way that increases the probability of both short-term and long-term success.

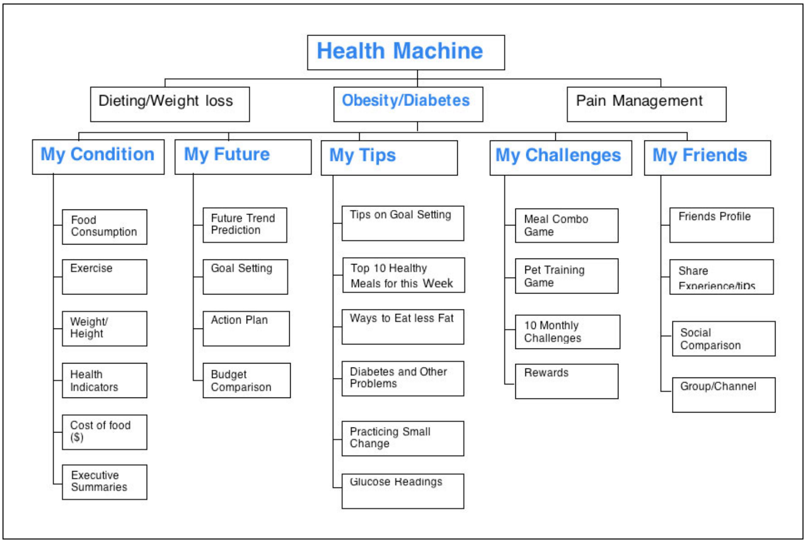


Fig. 1. Health Machine information architecture

Food Consumption. Records nutrition-components of food. Provides data-entry of information. Sums up the nutrition intake for each meal. Calculates and displays meal nutrition. Adds new food items to database and shares with other users. Provides ability to save selected food combinations for easy, efficient re-use.

Exercise. Records exercise type and duration. Sums up the energy consumption for each exercise activity**.** Outputs energy consumption. Adds new exercise items to database and shares with other users.

Weight/Height. Records weight/height data. Charts weight/height records and trends.

Health Matter indexes Monitoring. Because daily health-indexes monitoring is part of diabetes management, the Health Machine provides storing glucose readings. This enables the user to understand correlations between changes in diet and exercise and blood glucose levels. Users with diabetes also need to watch their heart rate and blood pressure. Users will be able to use the measures for setting goals and visualizing their process that may be of significance for their self-analysis purposes. The Health Machine also allows users to track their moods and associates them with their improvements of life-styles, a good incentive for behavior change. Compares the record with benchmarks and personal goals by table or chart display (*e.g.*, bar chart or line chart).

Food Cost. Records food expenses. Sums daily food expenses per meal components.

Health Thermometer. In the user-interface, the small health thermometer is always displayed at the top of the screen to show the user's current health condition summary. This display will be computed from a mixture of health factors that would be determined in consultation with a physician. This thermometer would be consulted as often as the user might monitor the mobile phone battery level or signal strength.

Executive Summaries. Sums up daily total calories, glucose levels, and blood pressure. Diabetes literature emphasizes testing glucose several times per day and at different times. Many people test glucose once per day, *e.g.,* in the morning before breakfast. Because exercise will reduce calories, and change glucose level and blood pressure, advanced Health-Machine functions might be able to calculate the calorie/glucose/BP levels in order to spare users having to do multiple tests every day. Outputs daily records.

My Future. Users will first view their current health scenario in a chart format. The chart illustrates users' current estimates based on current behavior: risk of heart attack, stroke, and diabetes in the next 20-30 years. Viewing these charts, users will have a visual impression of their health condition and the severity of their obesity and/or diabetes challenges. By changing the goal setting factors (weight, blood pressure, glucose, cholesterol level, *etc.*), users will view their different health scenarios in the future and thus decide the appropriate health indexes they would like to pursue and maintain. These series of health matter indexes or indicators will then be set as the users long-term goals. In accordance with the goals, suggested action plans will be provided. Users can customize their action plans, and then set them as their short-term plans. In this way, the users are expected to reach their long-term health goals though achieving the detailed action plans step-by-step. Their future health scenario will also serve as a critical incentive for them to keep going to achieve their longer-term goals.

Current Health Condition. Generates reports. Displays users' risk of having heart attack, stroke, hemiplegic paralysis, stupefaction, *etc.* Current data can be displayed by text, table, chart, visual image, map, or diagram, as appropriate.

Goal Setting. Offers a means for setting diabetes management goals. Set monthly health factors goals, such as weight, blood pressure, glucose, and cholesterol levels. Predict future heath condition change under different goals settings.

Action Plan. Generates recommended daily and weekly action plans based on monthly goals. Allows users to customize action plans.

Budget Comparison. Enables weekly food expense budget setting. Compares actual food expense with budget. Demonstrates cost savings for users with diet management and control.

My Tips. Provides tips on personal goal settings. Provides recommendations on healthy meals everyday. Updates tips through the Internet. Informs user about knowledge of complicating diseases resulting from obesity and diabetes. Provides tips about how to eat less fat. Provides advice how to increase exercise with little change in daily life. Top 10 Healthy Meals this Week synchs with mobile phone Website (or major diabetes/obesity research Websites). Updates top 10 healthy meals weekly. Ways to eat less Fat. Updates research articles on healthful, delicious diets and healthful cooking and eating styles. Diabetes and other Problems provides articles about diabetes and its complications. Practicing Small Changes provides tips about ways to increase exercise while incurring minor change in daily life. Glucose Readings provides articles about symptoms of low/high glucose levels and ways to deal with such conditions.

My Challenges. Apart from the tips, educates users through entertainment and games. For example, the Health Machine can offer games that can teach users to decide the right proportion and combination of food for their meals. In addition, a pet-training game could be designed for some users to have a better understanding of self-health management: how they treat a virtual "pet" reflects how they treat themselves. The 10 Monthly Challenges allows users to challenge themselves to achieve a series of behavior-changing goals. By accomplishing each of the challenges, users are expected to pursue short-term behavior adaptations and later form and maintain healthier life-styles. Virtual and real rewards would be provided to users, associated with winning games, competitions, and challenges. Users could share their reward information with their social groups, and gain more support from family, friends, and others. The Meal combo game teaches users to select wisely healthy food or food combinations for breakfast, lunch, dinner, and snacks. When players choose a healthier food, they will be rewarded a higher mark. They may also earn extra scores from healthy food combinations. Some people even give more care to their pets than to themselves. The virtual "pet" training idea is to track the exercise and food intake of the users and demonstrate the results in terms of how they treat their pets. Users will learn to behave healthier as they notice the change of the pet's mood and health condition. The pet will also send tips and notes to the users, and persuade them to eat healthier food and take more exercises.

My Friends. Adoption of a new, healthier lifestyle is best accomplished with the help of a support group. To encourage users to change behavior, the Health Machine would allow users to tap into social networks. Co-operation/competition within and among group members will serve as a strong social factor to motivate users' behavior change. Moreover, interactions can also consist of information-sharing among selected individuals through Facebook, Twitter, blogs, and other social media.

Friends Profile. Establishes friends profile database, and allows the participation of family members, friends, and celebrities.

Social Comparison. Enables individual comparisons on diet control, weight control, and exercise amount with friends, stars, people from other cities and countries, etc. Enables group competition. Friends and family members can form different groups and hold weight lost or exercise competitions. Users are expected to receive encouragements from group members and incentives to beat other teams.

Share Experience/tips. Users would be able to obtain more support from their social groups. For example: people having obesity or Type 2 diabetes problems can form social support groups to share experiences and advice with each other.

Group/Channel. Users can watch health, exercise, cooking, and food channels through their mobile phones to learn other related information.

8 Initial Sketches, Evaluation, Redesign, and Next Steps

Based on the information architecture, AM+A prepared initial sketches of key screens, which, together with a questionnaire, were intended to elicit feedback from potential users including healthcare trainers. AM+A planned to interview six people (two per persona) and two healthcare providers who train people who have been diagnosed recently with Type 2 diabetes. The survey for trainers appears in [Marcus, 2011]. Because of legal regulations related to official healthcare providers, it was possible to interview only one patient and one skilled trainer of patients. AM+A interviewed a Registered Clinical Dietician (RCD) from a major healthcare organization serving a wide demographic community in northern California. She provides training to obese and diabetic patients. Her detailed comments [Marcus, 2011] in particular,

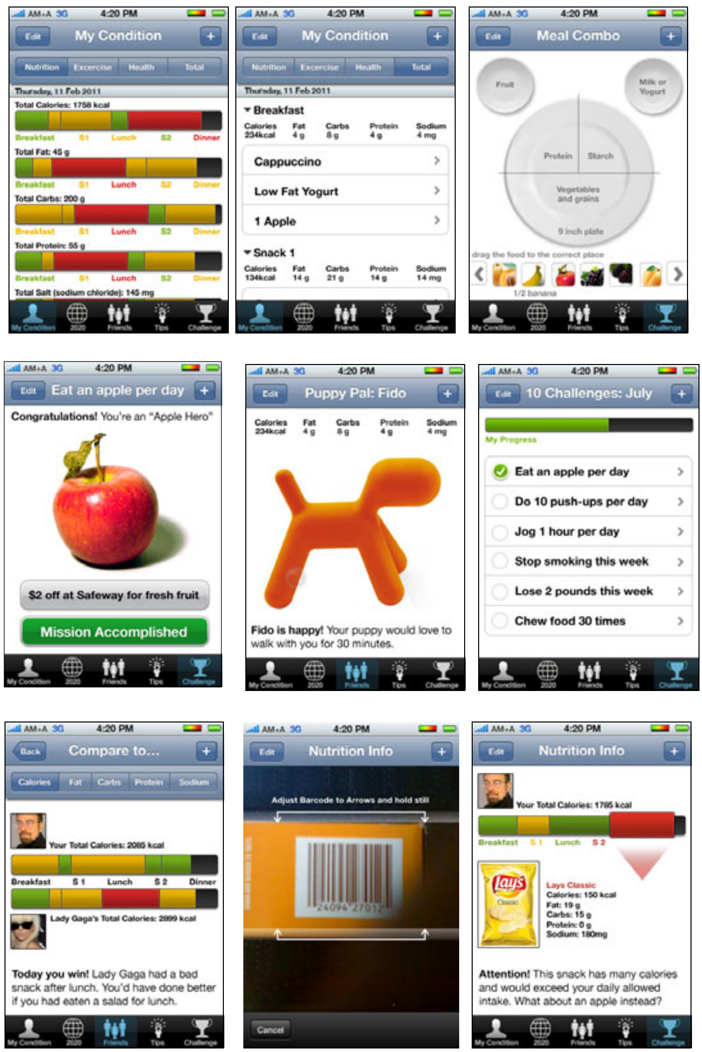


Fig. 2. Key Revised Screens

provided informally and anonymously, helped to drive changes in the screen designs. Based on the limited feedback, AM+A made minor changes in the information architecture, removing one item under My Condition. With regard to screen designs, AM+A made the controls more consistent with the platform guidelines (iPhone), and improved the simplicity of the text and imagery, in response to the nutritionist's specific comments. The accompanying figures display key screens.

Based on the user-centered design process described above, AM+A plans to continue improving the Health Machine. Tasks include: revise target personas and use scenarios; conduct multi-cultural evaluations; revise information architecture plus look and feel; and build initial working prototype. Future use scenarios will include desktop Website access in addition to mobile access to install or edit data.

9 Conclusions

The self-funded Health Machine project is a work in progress. AM+A has shared its approach and lessons learned. If the Health Machine approach is proven to be correct, it could have significant benefit for people in the US and elsewhere. AM+A is seeking to persuade other design, education, and medical groups to consider similar development. This incremental product development process has already been demonstrated successfully with a previous project, the Green Machine [Marcus and Jean, 2009], versions of which were taken up by a by major corporation for commercial development.

Acknowledgements. The author thanks AM+A Analyst/Designers Xuewei (Phoebe) He, Jue (Rosa) Yuan, Arne Berger, and Erika Reynolds for their significant assistance.

References

[1] Employees Earn Cash for Exercising More. Wall Street Journal 2, D3 (June 2010)
[2] Centers for Disease Control and Prevention (2008). National diabetes fact sheet: general information and national estimates on diabetes in the United States. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA (2007)
[3] Fogg, B.J., Eckles, D.: Mobile persuasion:20 perspectives on the future of behavior change. Persuasive Technology Lab, Stanford University, Palo Alto, CA (2007)
[4] Marcus, A.: The Health Machine: Information Design + Persuasion Design = Behavior Change Regarding Obesity + Diabetes. Information Design Journal 18(3) ((in press, 2011)
[5] Marcus, A., Jean, J.: Going Green at Home: the Green Machine. Information Design Journal 17(3), 233–243 (2009)
[6] Maslow, A.: A Theory of Human Motivation. Psych. Review 50, 370–396 (1943)
[7] Nat. Diabetes Info. Clearinghouse (June 2008). Nat. Diabetes Statistics (2007), `http://diabetes.niddk.nih.gov/dm/pubs/statistics/#prevention` (ret. July 05, 2010)

An Air Conditioning Control Method Based on Biological Fluctuation

Hiroki Matsumoto, Yoshio Iwai, Yutaka Nakamura, and Hiroshi Ishiguro

Graduate School of Engineering Science, Osaka University,
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
iwai@sys.es.osaka-u.ac.jp

Abstract. A living environment should be comfortable for all residents. The thermal environment is one of the indices of comfort, but it is difficult to adopt a specific thermal environment suitable for all residents who share a thermal space but have different personal needs. In this research, we propose an air conditioning control method that satisfies the demands of all residents in a given living environment. Control parameters change variously and control of the air conditioning is not easy for a typical real environment because the thermal environment changes dynamically. We propose a method of controlling air conditioners based on biological fluctuation.

Keywords: Air condition control, biological fluctuation, PMV, PPD.

1 Introduction

A living environment should be comfortable for all residents. The thermal environment is one of the indices of comfort in a living environment. However, it is difficult to adopt a specific thermal environment suitable for all residents who share a thermal space but have different personal needs. In other words, the demands on air conditioning in a living environment are different for each resident, and conventional air conditioners typically can only respond to one person's needs when responding to a remote control. Our research goal is to satisfy the thermal demands of all residents in a given living environment.

In this research, we propose an air conditioning control method that satisfies the demands of all residents in a given living environment. The proposed method controls the air condition using PMV values defined by ISO 7730 [1]. The PMV is an index that predicts the mean value of the votes of a large group of persons on the 7-point thermal sensation scale (+3 hot to -3 cold), based on the heat balance of the human body. It is desirable to control the PMV values, the temperature and the air velocities to bring the PMV close to zero.

J. Liang and R. Du have proposed an intelligent control system for thermal comfort using human learning and minimum power control strategies [2]. Their method, however, was tested using simulation data only and it is difficult to apply the method in a typical living space because the thermal environment of the living space changes dynamically and is affected by thermal noise. M. Mossolly et al. have proposed an air conditioning system using a genetic algorithm [3]. Their method controls the air

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 608–615, 2011.

temperature to a stable level in the whole room. Consequently their method cannot control the air condition for personal comfort because individual needs are different. F. Calvino et al. have proposed a control method for thermal comfort using a fuzzy adaptive controller. Their method has an adaptive property deduced from fuzzy logic. However, their method also focuses on the thermal environment and does not consider each person's individual needs [4].

For personal comfort, we must take into consideration each person's thermal state and control the air conditioners to respond to each individual's personal needs. However, the control parameters change variously and the control of the air conditioning is not easy for a typical real environment because the thermal environment changes dynamically. We propose a method of controlling air conditioners based on biological fluctuation.

2 Biological Fluctuation

A mathematical model for the adaptation mechanism of bacteria based on biological fluctuation has been proposed in [5]. This model is called the "attractor selection model" and is the simplest model for the control mechanisms of animals, using noises. Since this mechanism works flexibly and robustly without the model of the control target, we propose a control method based on this attractor selection model.

2.1 Attractor Selection Model

The attractor selection model can be expressed by the Langevin equation:

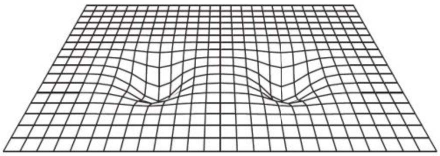
$$\dot{x} = -\frac{\mathrm{d}U(x)}{\mathrm{d}x}A + \eta, \quad (1)$$

where x and η are the state and the noise[1]. $U(x)$ is a potential function which has several attractors (local minima), as shown in Fig. 1 (a) and an attractor suited to the environment is searched for, by using noise. A (>0) is a called the "activity", which indicates the fitness of the state x to the environment. A also controls the behavior of the attractor selection model, i.e., $\frac{\mathrm{d}U(x)}{\mathrm{d}x} \times A$ becomes dominant in (1) when the activity A becomes larger, and the state transition becomes more deterministic. On the other hand, the noise η becomes dominant in (1) when the activity becomes smaller, and the state transition becomes more probabilistic.

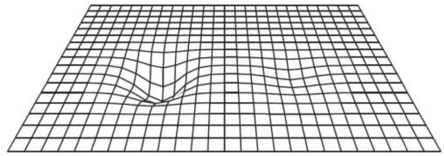
If the activity is large (small) in the left (right) half area in Fig. 1(a), the potential function is effectively modified as shown in Fig. 1(b). As a result, the state of a system is constrained into an attractor which is fitted to the environment where the activity becomes larger (the left attractor in Fig. 1(b)). Otherwise, the activity remains

[1] This formulation is not the only way to implement the attractor selection model, but it is convenient when explaining the behavior of the attractor selection model.

small and a suitable attractor is searched for by a random walk caused by the noise η .



(a) Potential function



(b) Effective landscape

Fig. 1. Attractor selection model: (a) and (b) show the potential field of attractors

2.2 Potential Function and Activity

It is difficult to design a potential function for a problem that is hard to model because the potential function is a mapping to the fitness from the state variables and depends on the dynamics of the target system. However, we can measure the degree of attainment of the task in some cases. For example, in the thermal environment we can use PMV values as the measure of thermal comfort.

In this research, we adopt the noise η according to the activity A to avoid the difficulty of constructing the potential functions of the target system.

3 Air Control Method

3.1 Air Control Method Based on Biological Fluctuation

The activity of each air conditioner $A_j(t)$ is calculated using the following equation:

$$A_j(t) = \sum_i w_{i,j} a_i(t) \quad , \tag{2}$$

where $a_i(t)$ is the activity of the PMV sensor described in a later section, and control parameter $x(t)$ is update by

$$x_j(t) = x_j(t-1) + \eta_j \, . \tag{3}$$

In this experiment, the components of $x_j(t)$ are the air temperature, air volume, and air direction of air conditioner j . Noise is given by η_j which follows $N\left(0, \sigma_j^2(t)\right)$. The activity is used to calculate $\sigma_j^2(t)$ as follows:

$$\sigma_j^2(t) = u \exp\left(-A_j(t) s\right) , \tag{4}$$

where u and s are constant. From these equations, when the activity becomes larger, the fluctuation of the control parameters becomes smaller, and vice versa.

3.2 Design of the Activity of PMV Sensors

The activity of each air conditioner $A_i(t)$ is calculated as the weighted sum of the activities of the PMV sensors, $a_j(t)$ as shown in Equation (2). The weight w_{ij} expresses the effect of the *i*-th air conditioner to the *j*-th PMV sensors. In a typical real environment the weights are dynamically changed because of environmental changes such as humans, moved objects, and so on and it is hard to estimate the true value of the weights. Such an uncertainty can be resolved by optimization based on biological fluctuation.

The activity of the PMV sensor, $a_j(t)$ becomes small when the value of the PMV sensor, $p_j(t)$, is nearly equal to zero, and vice versa. The response of the PMV sensors after the change in the parameters of the air conditioners is very slow when the power of the air conditioners is small relative to the space in a living room. Therefore, we should incorporate the gradient information of the PMV sensor, $\Delta p_j(t) = |p_j(t)| - |p_j(t-1)|$, in the calculation of the activity of the PMV sensors. In this paper, we use Table 1 to determine the activity.

Table 1. Determination rule for the activities of the PMV sensors

$\Delta p_j(t)$ \ $\lvert p_j(t)\rvert$	$\lvert p_j(t)\rvert < 0.3$	$0.3 \le \lvert p_j(t)\rvert \le 0.7$	$0.7 < \lvert p_j(t)\rvert$
$\Delta p_j(t) > 0.0$	0.0	0.0	0.5
$0.0 \ge \Delta p_j(t) > -0.2$	0.0	0.5	1.0
$-0.2 \ge \Delta p_j(t) > -0.5$	0.5	1.0	0.7
$-0.5 \ge \Delta p_j(t)$	1.0	1.0	0.5

3.3 Avoiding the Selection of Incorrect Parameters

When all the values of the PMV sensors are positive, the air conditioners should be controlled to cool a living space and should not be controlled to warm it, and when all the values of the PMV sensors are negative, the air conditioners should be controlled to warm a living space and should not be controlled to cool it. In such a simple control case, it is easy to find that the control parameters are incorrect, so we recalculate the control parameters using Equation (3).

4 Experimental Results

We conducted experiments in an experimental living room to show the effectiveness of the proposed method. The experimental living room is shown in Fig. 2 and the layout of the room in Fig. 3. There were three air conditioners and two PMV sensors in the living room. The air temperature, mean radiant temperature, humidity and air velocities were measured by the PMV sensors. In the experiments, clothing insulation was fixed at the same value and the metabolic rates were changed in the experiments as shown in Fig. 4. The PMV values are calculated from these values.

Fig. 2. Experimental living room

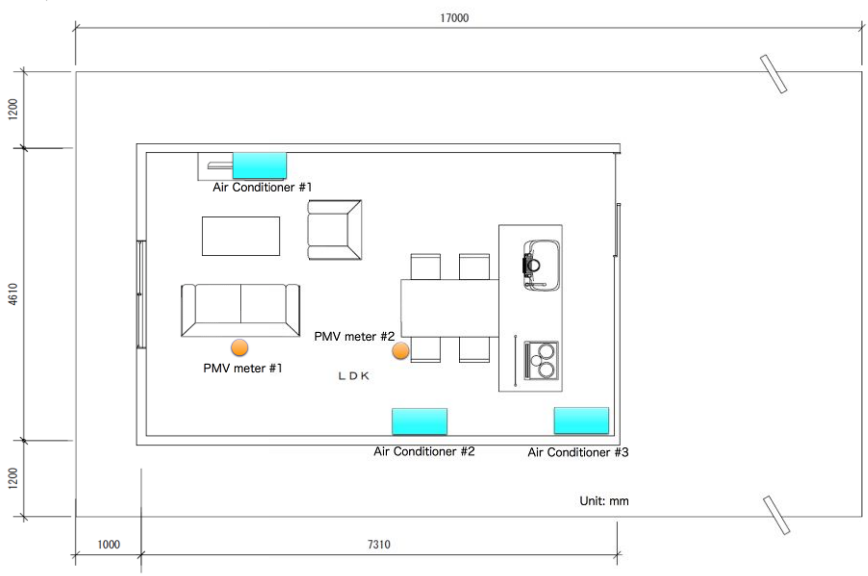


Fig. 3. Layout of the experimental living room: three air conditioners are shown by rectangles and two PMV sensors are shown by circles

We compared the results of the three control methods: 1) built-in automatic control in the air conditioners, 2) manual control by a person in the living room, and 3) automatic control by the proposed method. The number of test subjects was 6 and each controlled the air conditioners through GUI applications by monitoring two PMV values measured in real time.

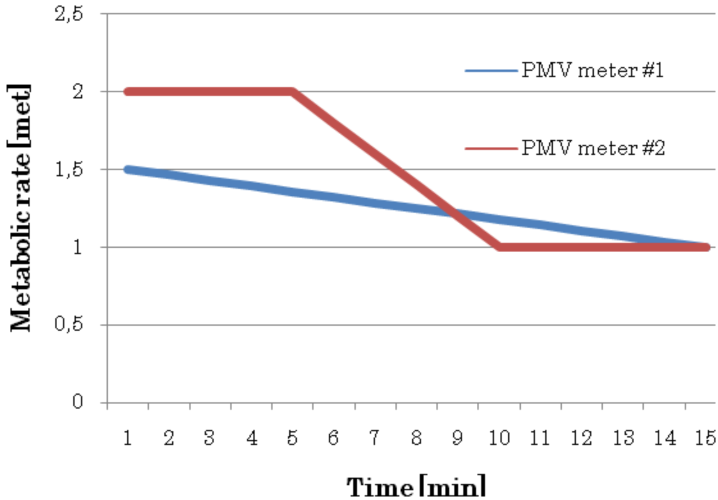


Fig. 4. Metabolic rates are changed through the experiments

Examples of the experimental results are shown in Figs. 5, 6 and 7. We define three criteria for evaluation: a) the average of the sum of the absolute values of the PMV sensors, b) the ratio of the time that the absolute values of the PMV sensors were greater than 0.5, and c) the area that the absolute values of the PMV sensors were greater than 0.5. When these values become smaller, the probability that a person feels comfortable becomes higher. Table 2 shows the values of these criteria in the experiments. From the table, the proposed method can control air conditions to a comfortable state and the method has almost the same performance as manual control.

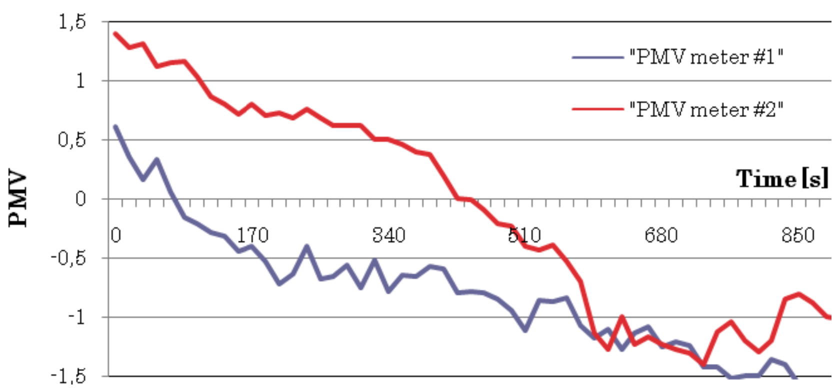


Fig. 5. An example of the PMV values controlled by built-in automatic air controller

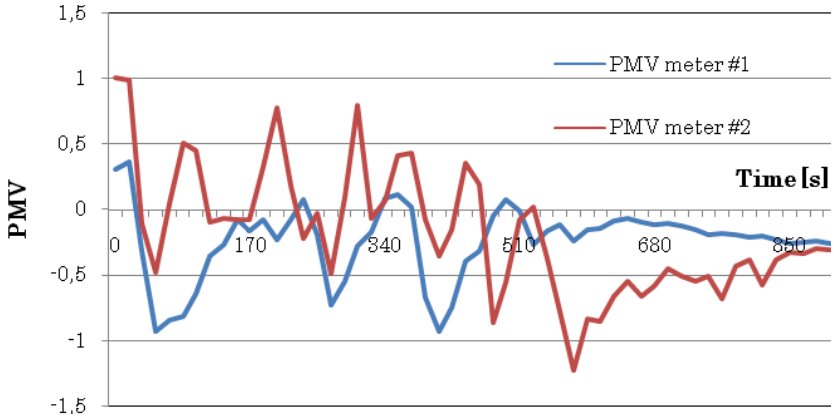


Fig. 6. An example of the PMV values manually controlled by a person

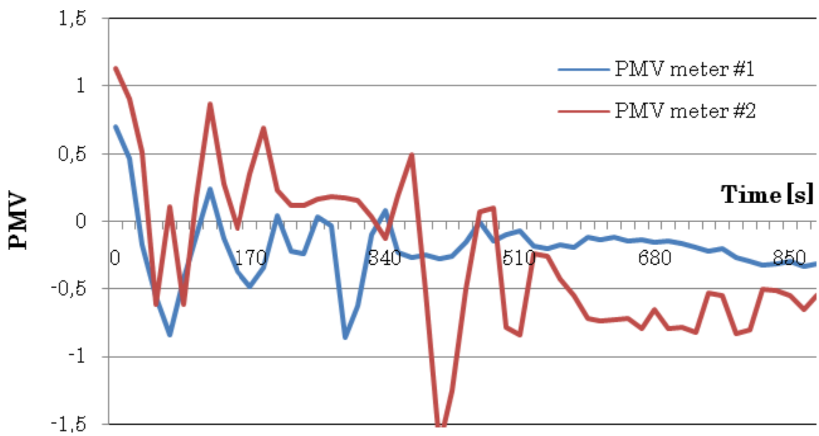


Fig. 7. An example of the PMV values controlled by the proposed method

Table 2. Comparison of the control results

Criteria [unit]	Automatic	Manual control	Proposed method
a [PMV]	1.70	0.88	**0.72**
b [%]	0.60	0.35	**0.23**
c [PMV•s]	43	25	**6.3**

5 Conclusions

In this paper, we have proposed a method for the control of air conditioning by imitating a biological fluctuation. Conventional methods require much information including the 3D room shape, air flows, and distribution of temperature for precise control.

However, in a typical real life situation conventional methods cannot control air conditions precisely because some parameters are difficult to estimate and contain much noise. Though the proposed method does not take into account such parameters, the method can control the air condition in a stable state from the experimental results as an imitation of a biological fluctuation.

The drawback of the proposed method is that PMV sensors are required for activation control of the biological fluctuation. In future work, we propose to replace the PMV sensors with other cheaper sensors such mobile phone cameras, IR sensors and so on.

Acknowledgments. This work is partially supported by cooperative research with the Daiwa House Industry Co., Ltd.

References

1. ISO: Ergonomics of the thermal environment — Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, ISO-7730, 3rd edn. (2005)
2. Liang, J., Du, R.: Design of intelligent comfort control system with human learning and minimum power control strategies. International Journal of Energy Conversion and Management 49(4), 517–528 (2008)
3. Mossolly, M., Ghali, G.N.: Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm. International Journal of Energy 34(1), 58–66 (2009)
4. Calvino, F., Gennusa, M.L., Rizzo, G., Scaccianoce, G.: The control of indoor thermal comfort conditions: introducing a fuzzy adaptive controller. International Journal of Energy and Buildings 36(2), 97–102 (2004)
5. Kashiwagi, A., Urabe, I., Kaneko, K., Yomo, T.: Adaptive response of a gene network to environment changes by fitness-induced attractor selection. PLoS ONE 1(1), e49 (2006)

First Validation of Persuasive Criteria for Designing and Evaluating the Social Influence of User Interfaces: Justification of a Guideline

Alexandra Némery[1,2], Eric Brangier[1] and Steve Kopp[2]

[1] Université Paul Verlaine – Metz. InterPsy-ETIC, EA 4432. User Experience Lab., Faculté des Sciences Humaines et Arts BP 30309 Île du Saulcy - F-57006 Metz, France
brangier@univ-metz.fr
[2] SAP-BusinessObjects Division, 157-159, rue Anatole France. F-92309 Levallois-Perret, France
{alexandra.nemery,steve.kopp}@sap.com

Abstract. Ergonomics has often produced grids to measure the ergonomic quality of goods and services. This paper seeks to establish and validate a grid to focus on the persuasive dimensions of interfaces and their effects; a grid that is robust, reliable, useful, relevant and easy to use for ergonomists, usability engineers and interaction designers. Our purpose is to develop and validate guidelines to measure and assess the persuasive dimensions of user experiences. This research based on a criteria model will become helpful in researching and designing persuasive technology. At first we propose a criteria-based approach to measure persuasive strength of interfaces; this criteria grid includes eight criteria: Credibility, Privacy, Personalization, Attractiveness, Solicitation, Priming, Commitment, and Ascendency. At the end, these criteria are validated by a sample of 30 experts, who confirmed that the proposed categorization of the criteria (from 71.30% to 83.25%).

Keywords: Persuasive technology, Criteria grid, Captology, Ergonomics' guidelines, Media and social influence.

1 Introduction

We are shaping technologies and at the same time technologies are shaping us. They influence us and tend increasingly to modify our behavior. Technologies are increasingly becoming social actors [2]. In organizations, software is used for an increasing range of collaborative activities, sometimes including marketing and consumer research. Consequently, a lot of research is conducted in the area of persuasive technology; persuasion is beginning to invade our technical systems, especially in the area of social networks, e-commerce, e-health or e-learning. The purpose of this chapter is to define and validate criteria to design and assess persuasion in HCI (Human-Computer-Interaction). Our objective is to propose a framework for the analysis of persuasive elements as a guideline for the design of interfaces.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 616–624, 2011.

To build this guideline, we relied on assessment criteria from a demonstrated theoretical basis [1, 7, 8, 10]. Theoretical background must represent a good foundation for a relevant and consistent architecture [10]. Beyond these aspects, we need models and criteria gathering all the important aspects of persuasion but also with a high level of comprehensiveness [11, 12]. We also need research to validate criteria.

In this paper, we will describe an experiment conducted with usability specialists, user experience designers, ergonomists and human factor engineers to evaluate the relevance of a structured guideline for designing and assessing the persuasive dimensions of human-computer interfaces. These guidelines could be applicable to different types of devices (website, software, game, mobile) and to wide range of applications. We will first present the theoretical background and the conceptual framework of the proposed criteria for evaluating and designing persuasive interface. Then, we will explain methodology developed to validate the persuasive criteria approach. Finally, we will analyze our results and give comments on relevance, interests, limitations and the possible refinement of this persuasive criteria guideline.

2 Usefulness of Guidelines for Designing and Evaluating Persuasive Interfaces

2.1 Brief History

Initially, computers were simply tools with great capabilities of storage and calculation of huge volume of data. Their user interactivity was limited, and there were no attempts to use them as an influence appeal. The first cases for using the persuasive power of technology took place in the 1970s. It aimed at promoting health related safe behavior in workplaces and to improve employee productivity. These studies continued during the 1980s, but the potential of persuasive technologies really started to be exploited in the 1990s with growth of successful e-Shopping sites. They invade the Web and become great success with a new competitive and innovative approach. These techniques are deployed to exploit the advantages of this media like interactivity and attractiveness. Obviously, the aim of websites creators is to increase consumer behavior.

2.2 Technology and Social Influence

Fogg [5] opened the way for the new field of persuasive technology for which he coined term captology, an acronym for “computers as persuasive technologies”. The concept of persuasion can cover a range of meanings, but Fogg defines it as “an attempt to shape, reinforce, or change behaviors, feelings, or thoughts about an issue, object, or action”. Fogg [5] presents persuasion technology as both (a) a tool, since technology can help people achieve their objective, (b) as a media interaction which creates an experience between the user and the product, and (c) as a social actor, which means that technology deals with the problem of strategies for social influence and compliance.

Persuasive technology is a medium to influence and persuade people through HCI. Indeed, technology becomes persuasive when people give it qualities and attributes that may increase its perceived credibility, privacy, personalization and attractiveness, for example. Persuasion in HCI is at the crossroads between ergonomics, social psychology, behavioral economics, organizational management and obviously the design of user experience.

Persuasive technology has been applied to many domains. Technology development also initiates a diversification of applications. The rise of e-shopping websites in recent years is propitious to the use of persuasive methods, both in the field of design and ergonomics, trying to change purchasing behavior. This explains why marketing is a beacon for persuasion technology. eBay", a pioneer eCommerce web site, devotes its success at least partially to persuasive techniques like the seller's star ratings, which provides buyers with increased confidence that a transaction will be fulfilled. A major new field of research concerns the field of health, both in the prevention of risk, monitoring of disease, and the promotion of sport [2].

To sum up: all areas of our life can be affected by persuasion technology whether in education, health, consumption, entertainment and especially in the workplace. These elements allow us to say that it is required to develop ergonomic practices and to integrate this concept during software design process.

2.3 Importance of Guidelines to Evaluate Persuasion in HCI

The idea to evaluate persuasion technology is not well accepted because it is judged to be unusual, antisocial, time-consuming, and useless to explain to users that the computer manipulates them, etc. Over the last decade, the critical mass of research of this topic [3, 11] clearly shows the importance of the field. One recommendation for future research is to create methods for a clearer measurement of persuasive system, like cognitive walkthrough method or heuristic evaluation, specially built for the evaluation of persuasion in HCI.

The aims of the guidelines methods are to intervene early in the interface design procedure, to categorize, identify, meet the requirements and quantify usability problems dealing with persuasiveness, and finally to be integrated into a design life cycle. Guidelines give heuristics, which are stabilized knowledge on persuasion that experts in usability could use during the evaluation and design processes; usability specialists compare each element of the interfaces with principles and heuristics coming from persuasion concepts and theories.

The guidelines purpose is therefore to measure the persuasive dimensions involved in usability, to help experts to test the interface using a check-list. As this method requires a high degree of expertise, our first work [8, 9] was to complete, synthesize and establish criteria.

3 Criteria for the Assessment of Technological Persuasion

Computer ergonomics has yielded a number of criteria grids to help with the measurement of the ergonomic quality of goods and services [1]. We seek to establish a tool to focus on the persuasive dimensions of interfaces and their effects; a grid that is robust, reliable, useful, relevant, and easy to use. Our proposal is based on a bibliographic analysis and defines a grid that distinguishes forms and processes of social influence, respectively the static and dynamic aspects of the interface (tbl. 1).

Table 1. General architecture of guidelines in persuasive human-computer-interaction

Ergonomics criteria for persuasive interactions: Eight criteria to measure technological persuasions in human-computer interface	
b	b
Static criteria In interfaces, some prerequisites are necessary to promote the acceptance of an engaging process. These criteria are based on the content of technological influence.	**Dynamic criteria** Regarding dynamics, there is also a means to bring the user in a process of interaction to strengthen the progressive engagement of the user to the elements of its interface.

b	b	b	b	b	b	b	b
Credibility of Interaction	**Guarantee of Privacy**	**Personalization**	**Attractiveness**	**Solicitation**	**Priming**	**Commitment**	**Ascendency and Possibility of Addiction**

Most guidelines or inspection methods are intended for use by experts, so their contents must be defined (tbl. 2) and, obviously, validated; even if the process of persuasion has to be carefully analyzed both in given situations as well as in a larger technological context.

Table 2. Description of the 8 persuasive criteria

Criteria	Definition	Justification	Example
Credibility of Interaction	Giving enough information to the user enables him to identify the source of information to be reliable, expert and trustworthy.	Credibility affects use and is seen as a form of loyalty. Credibility is the combination of the perceived reliability and perceived expertise of the product.	Presenting updated information and the date of the update.
Guarantee of Privacy	Do not persuade the user to do something that publicly exposes his private life and which he would not consent to do.	Privacy is an important aspect of ethics. Respect for privacy is one of the most dominant concerns about people's behavior with internet communication.	The system should preserve wherever possible the right of a user to remain anonymous to the larger user community as well as the providers of the system.

Table 2. (*Continued*)

Personalization	Present information adapted to the user or to the user group.	Customization makes the message more relevant and the information will be integrated by the user more quickly and will better draw his attention.	Remembering the state from a user's last interactions
Attractiveness	Presentation of elements in a way that is engaging and visually appealing.	Enhancing surface is related to the persuasive design. Controlling the physical elements of the interface and maximizing the visual impact can lead to increased loyalty and create or reinforce a behavior.	The choice of colors as a reinforcement of the message.
Solicitation	To solicit user in a light way, to catch his attention with an engaging effect.	Using the information given by the user allows for opportunities to engage the user.	Display many invitations.
Priming	Trigger user interaction by creating a point of entry, stirring interest.	Priming enables the user's initial interest and interactions	Click to immediately view relevant and interesting content for free
Commitment	Reinforce repeat behavior and frequency, as well as individual engagement.	Having engaged in an inexpensive first step, it will be easier to accept the following steps, each time increasing the persuasive influence.	Improve the frequency of the final behavior or attitude expected.
Ascendency and Possibility of Addiction	Show engaging scenario completion, follow up its influence and control its evolution over time.	The last step is the culmination of the process leading to behavior and attitude initially expected. We can then speak of voluntary submission.	The individual accepts information that he would not have accepted voluntarily.

4 Method of Validation

The aim of the study was to assess the effectiveness of our grid by using the methodology based on the comprehension and the categorization of the purposed criteria by experts in user experience.

Our grid was explained to 30 usability specialists. On average, participants had 13 years of experience (SD=7,9) in the evaluation and/or design of HCI. There were 10 women and 20 men. Each criterion was presented in a document with a definition, a justification, references and illustrated examples (pictures, screens, messages, etc.). Some of them were subdivided into sub criteria but the experiment was only on the 8 elementary criteria. A total of 15 interfaces from Internet, Smartphone and video games were used. 84 elements of these interfaces have been identified by multi expert judgment as persuasive elements. Each subject participated in individual experimental session.

The first phase of experiment consisted in learning the 8 criteria. Individuals were invited to read the definition document, the justification and were encouraged to give special attention to examples. In the second phase, participants were asked to evaluate 15 interfaces from commercial, education, ecology, business, social network and entertainment field and randomly presented. The participants were allowed to consult the document with description of each criterion but without any examples. The material tested was static interfaces. Limitation is due to the fact that the inspection was on non interactive system.

The score of correct identification was calculated for each criterion.

5 Results Analysis: The Relevance of the Persuasive Criteria

The mission of the evaluators was to note and comment any persuasive aspects of the 15 interfaces occurring during the use of our grid and to judge the method. They suggest the following analysis.

The global score of correct identification is 78.8% (fig. 3). Identification was considered correct if it was in line with the experts during the pre-test.

The results show that using the grid makes it possible especially to detect problems linked to the persuasiveness of multimedia interfaces; the problems linked to the dynamics criteria (long time effects) are more difficult to define, in particular the addictive process.

From this study, the results indicate that the criterion that should be improved is the last dynamic criterion: Ascendency. In fact, the last dynamic criterion is the achievement of a persuasive process. Most of the time it could not be shown on a HCI but only observed in real life (like stop smoking for health website) or interpreted by some elements on the screen (like time overconsumption in video game in user profile page.) Some superficial modifications are needed to other definitions, especially on wording aspects and by providing more typical examples to illustrate each criterion. This would remove the issue about heuristic evaluation being unable to measure the “ascendency” towards a final goal.

In addition to our results, we will discuss the fact that the acquisition of new concept from marketing, design and, even social psychology are not simple. It was difficult to allow the participant to understand the dynamic process by non interactive and static screens. Despite the improvements needed (Ascendency criteria), the results are very encouraging.

Table 3. Correct identification table

Static criteria (83.25%)	Credibility	81.5%
	Privacy	90,7%
	Personalisation	82.9%
	Attractiveness	77.9%
Dynamic criteria (71.30%)	Solicitation	75.3%
	Priming	75.4%
	Commitment	76.7%
	Ascendency-Addiction	57.8%

The validation is based on an expert's method evaluation. In this case, the number of experts is fairly high (n=30): usability specialists, ergonomists, interface designers. Besides this, in the majority of cases when an interface is designed, the potential users are relatively well-known: the method is applied strictly and systematically. In real multimedia projects, the knowledge of the user is often limited and the diversity of the public targeted greatly hinders the generalization of reasoning on the potential defects in persuasive interfaces. In particular, the estimation of the degree of severity of "Ascendency-Addiction" proved to be more or less impossible for some evaluators.

That said, all persuasive aspects could not be reduced to a criteria set. It requires more knowledge on the task, the use context and more information on user specificities. In short, our grid is one inspection methods with which an expert analyses and criticizes the persuasive dimensions of an interactive system to be evaluated.

6 Conclusion

The main aim of interactivity is to provide the users with a coherent interaction structure that is to say to enable them to accomplish their goals using a path that is both efficient and satisfactory. During this interaction process, users are often influenced by technological frames which try to manipulate their opinions, their choices, and a satisfactory level of compliance. Persuasive technologies try to provoke interest and action. The measurement of persuasion in HCI is, consequently, critically important for understanding how persuasive design could influence user goals and expected outcomes, in a way that removes both real and perceived obstacles and enables a closer match with user goals and outcomes expected by a system's designers.

The intention of this study based on an experimental identification task was to evaluate the definition of our criteria grid proposal used to help experts to better evaluate or design persuasive interface. Our evaluation guideline aims at covering what we call the persuasive strength of interactive system. This strength could be described in 8 elementary criteria: credibility, privacy, personalisation, attractiveness, solicitation, priming, commitment, and ascendency. The evaluation of this guideline shows the criteria considered to be relevant to inspection methods usable in persuasive interfaces.

Persuasive criteria cover all aspects that try or aim to influence users. It takes into account all the dialogue form that the interaction between user and software could have. It tries to be predictive in the sense that it helps to diagnose a probable attitude or behaviour. These criteria could also serve as guidelines to guide the choice of interaction design, graphic design, electronic messages, and user experience as well. The performance of inspection tools must take into account user profile and situation diversity and this method must surely be supplemented by other evaluation methods but results of this experiment are encouraging to keep on validation of the grid in other ways.

Usability inspection is the generic name for methods that inspect the usability of the user interface. In this paper, we showed that a grid for evaluating persuasion in HCI is not only interesting, but useful to design and evaluate user experience. We also underline that a validation of persuasive criteria could be done. Especially, we present a guideline and show that our eight criteria could be used for evaluating the ergonomic quality of user interfaces. The criteria help experts to identify more persuasive elements in interfaces. The design of an interactive system is an iterative process, and evaluation is an imperative component of this development. For this reason, evaluation has been the subject of numerous studies and much research in the scientific community concerned with HCI [6]. Many evaluation methods and techniques exist but persuasion is out of the scope of inspection. This research should help to integrate persuasive criteria to usability inspection tools.

Acknowledgments. The authors wish to thank Michael Thompson, Director of Product Management at SAP, for his numerous insightful comments on previous versions of the paper.

References

1. Brangier, E.: Besoin et interface. In: Akoka, J., Comyn-Wattiau, I. (eds.) Encyclopédie de l'informatique et des systèmes d'information, pp. 1070–1084. Vuibert, Paris (2007)
2. Brangier, E., Bastien, J.-M.-C.: Ergonomie des produits informatiques: faciliter l'expérience utilisateur en s'appuyant sur les notions d'accessibilité, utilisabilité, émotionnalité et d'influençabilité. In: Vallery, G., Zouinar, M., Leport, M.-C. (eds.) Ergonomie, conception, de produits et services médiatisés, PUF, pp. 307–328 (2010)
3. Chak, A.: Guiding Users with Persuasive Design: An Interview with Andrew Chak, by Christine Perfetti, User Interface Engineering (2003)
4. Dhamija, R., Tygar, J.D., Hearst, M.: Why Phishing Works. In: Proceedings of the Conference on Human Factors in Computing Systems, CHI 2006, Montréal, Québec, Canada, April 22-27 (2006)
5. Fogg, B.J.: Persuasive Technology. Using computers to change what we think and do. Morgan Kaufmann, Amsterdam (2003)
6. Mahatody, T., Sagar, M., Kolski, C.: State of the Art on the Cognitive Walkthrough method, its variants and evolutions. International Journal of Human-Computer Interaction 26(8), 741–785 (2010)
7. Nemery, A., Brangier, E., Kopp, S.: Proposition d'une grille de critères d'analyses ergonomiques des formes de persuasion interactive. In: David, B., Noirhomme, M., Tricot, A. (eds.) Proceedings of IHM 2010. International Conference Proceedings Series, pp. 153–156. ACM, New-York (2010)
8. Nemery, A., Brangier, E., Kopp, S.: How cognitive ergonomics can deal with the problem of persuasive interfaces: Is a criteria-based approach possible? In: Norros, L., Koskinen,

H., Salo, L., Savioja, P. (eds.) Designing beyond the product: understanding activity an user experience in ubiquitous environments, European Conference on Cognitive Ergonomics, ECCE 2009, pp. 61–64 (2009)
9. Nemery, A., Brangier, E., Kopp, S.: Élaboration d'une grille d'analyse des aspects persuasifs en ergonomie informatique. Actes du 44 congrès de la Société d'Ergonomie de Langue Française: Ergonomie et organisation du travail. Toulouse, pp. 483–490 (2009)
10. Nielsen, J.: Heuristic Evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods. John Wiley & Sons, Inc., New York (1994)
11. Oinas-Kukkonen, H., Harjumaa, M.: A Systematic Framework for Designing and Evaluating Persuasive Systems. In: Oinas-Kukkonen, H., Hasle, P., Harjumaa, M., Segerståhl, K., Øhrstrøm, P. (eds.) PERSUASIVE 2008. LNCS, vol. 5033, pp. 164–176. Springer, Heidelberg (2008)
12. Oinas-Kukkonen, H., Harjumaa, M.: Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems 24, article 28, 485–500 (2009)

Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness

Tanner Olsen, Katelyn Procci, and Clint Bowers

Department of Psychology, University of Central Florida,
Orlando, FL, 32816, USA
{t_olsen,kprocci}@knights.ucf.edu, bowers@mail.ucf.edu

Abstract. Usability testing is an important, yet often overlooked, aspect of serious game development. Issues in usability can drastically impact user experience and thus the learning outcomes associated with serious games. The goal of this paper is to provide serious game developers with an approach to efficiently and effectively apply usability testing into their development process. We propose a three-tiered approach to the assessment of game usability with the addition of assessments playability and learning to traditional usability. Learning or training is the main objective of a serious game and enjoyment is often required when trying to elicit the necessary usage to achieve this goal. Step-by-step procedures and associated measures are provided to assess usability, playability, and learning outcomes concurrently with game development, while taking into account the unique goals and limitations of time, personnel, and budget that small development companies often encounter.

Keywords: usability, user experience, serious games.

1 Introduction

Usability is one of the central elements in the game development process. It is deeply rooted in the overall experience of the player and can affect their interaction with the game. For example, if a player cannot read the text on screen or the controls are difficult to master or are unresponsive, the failure of ensuring good usability detracts from the overall experience. Serious games in particular present unique challenges in regards to effective usability testing. Briefly, serious games have been introduced into a growing number of domains including education, therapy, and personnel training, with the goal of providing a method to supplement traditional means of learning. Accomplishing this goal requires converting the tenets of effective training and learning into game features while utilizing a usable game design. If overall usability fails and all of the player's effort is put toward mastering controls, not much attention and cognitive reserve remains to focus on the actual game content.

In small serious games development laboratories, resources are often extremely limited. Constrained by budget, time, and small development teams, usability in serious games often comes as an afterthought. Serious game development is also unique in that it often incorporates researchers who are interested in the science of learning

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 625–634, 2011.

and games. This places further constraints on development as, during the usability process, you need to address two very distinct audiences with two very different needs while creating a singular product. The aim of usability testing, then, is to give meaningful and immediate feedback to the developers while providing useful data for researchers.

We have created a procedure for usability testing that addresses the needs of both the developer and the researcher. Our procedure is different from other usability approaches because it addresses the needs of the developer while providing quantitative data for the usability expert. It also addresses the needs unique to serious games, namely usability, playability, and learnability.

1.1 Usability, Playability, and Learnability

There is a dynamic interaction of closely related, yet ultimately unique, components that affect the success of the serious game as a training tool. Not only must you assess the game for basic usability, playability and educational merit are also critical. For example, "user interface must not merely be functional or easy to use - it must also be fun!" states game designer Chris Crawford, "if the game interface is clumsy or confusing, the player simply abandons [the game] [1]."

There have been a number of scales and measurement techniques developed to assess the usability of any number of distinct systems. Computer systems and programs have been the focus of many usability scales due to their importance in modern society, so there is no shortage of scales available for assessment of various features. Often, many technology-based companies design their own usability assessments in order to create scales relevant to their particular areas of development. In addition, standardized and validated scales can be used. These scales generally consist of Likert scales that focus on different aspects of usability such as display characteristics, including location of information on screen and legibility; language usage; the ease of interaction with the program and difficulty carrying out desired tasks; how easily the system is learned; general consistency; and other subjective measures associated with how well the system operates. These scales range in length from only several to hundreds of questions of varying specificity.

Usability is a more micro approach, focusing on the independent functionalities within individual components of a system. Playability, on the other hand, focuses on a broader sense of overall functionality associated with the integration of several usable tools, allowing for successful and enjoyable interaction with a game. Resnick and Sherer defined play as "entertainment without fear of present or future consequences; it is fun" [2]. For example, if usability testing found that control mechanics for moving a player character and interacting with in-game objects were excellent, yet playtesters are unable to integrate the two functions at the same time, the game is rendered frustrating and unplayable rather than engaging and enjoyable. Playability is therefore a more elusive component to capture. As a holistic experience, playability has been desired in serious games, but there are not any widely used measures for it. There are, however, associated measures that share similar components to those that are of interest in serious game usability testing: these include scales of immersion and presence [3], flow [4], and engagement [5]. Such scales have largely been developed to study the effects of simulations as well as entertainment-based video games on individuals, but also can provide an excellent starting point for assessing playability within the context of serious games.

Finally, it is essential that learning outcomes be measured for individuals during the testing phases of the game to ensure that the game achieves its primary objective before too much time, effort, and budget have been committed to the project. Focusing too much attention on adopting the characteristics of the entertainment-based counterparts of a game can result in the sacrifice of learning effectiveness. Poor usability can also impair learning by taxing cognitive resources and decreasing motivation to use the game. Therefore, assessing learning outcomes at various stages during development can help determine what might be the cause of increases and decreases in learning, and help the development team maintain optimal focus on the most important feature of the serious game.

1.2 Functional Balance of Usability Components

It is important to maintain a functional balance of each component to promote an optimized level of learning and desire of use. Unfortunately, as mentioned previously, comprehensive usability in serious games is often applied as an afterthought late in the process due to financial constraints and restrictive deadlines. As shortcomings in any of these components can potentially undermine the others, the goals of the serious game will also likely be compromised. As it currently stands, this design approach is ineffective and should be addressed. Though previously developed and validated scales can be very useful to conducting usability testing of your own, pre-made tests can also be limited in their scope. Serious games pose some unique challenges when trying to adopt current scales because they are unlike typical programs subjected to usability testing. It is often necessary to adopt and adapt multiple scales toward measuring each important aspect of the game including the usability, playability, and learning outcome.

2 A Comprehensive Strategy for Usability in Serious Games

As a serious games research organization, we strive to uncover the best practices for both design and development. This involves providing practical guidelines for programmers and game designers while cultivating quantitative tools for analysis by usability professionals and researchers. Thus, it is our goal to present developers with a technique to assess usability, playability, and effectiveness that can be applied directly to their own game development cycles.

We propose a three-tiered approach to be applied throughout the development process to maximize the effectiveness of serious games in which classic usability, playability, and educational merit are analyzed at several intervals in order to guarantee that these important pieces are not sacrificed at any stage of development. This approach has been designed with small developers in mind and has been optimized to maximize return on investment (ROI). We will provide extensive resources and guidelines that can be used to improve your own internal development models to ensure the development of functional, enjoyable, and effective end products.

2.1 Pre-development

Prior to serious game development, it is important to properly identify the target audience and the user characteristics. Subject matter experts (SMEs) can be useful when compiling a user profile, but it is important to also contact and interview within the target population to assure that potentially important details are not overlooked as SMEs will not know everything.

Age, gender, and basic demographic information should be collected as well as any further background information that may influence the interaction with the software. This additional information includes previous knowledge and experience with the material to be covered, gaming experience, reading level, and other user capabilities and limitations, such as disabilities, that may impact interaction with the game. If the background information desired is highly sensitive (religious views, sexual history, etc.), resulting in participants refusing to respond to items, it may be useful to place this survey at the end of the document, as this may increase the response rate due to the amount of commitment the respondent has already made to the study [6] [7].

A cognitive task analysis should be conducted to determine the demands of the task and designs should be implemented that reduce the effort exerted towards secondary tasks such as controls and menu navigation. Also, baseline knowledge in this area should be assessed prior to game development to determine what material is necessary to cover and the amount of time that should be dedicated to each section of the material. Measuring the perceived relevance of the proposed serious game before production is also greatly beneficial. This allows the developer to determine whether the target audience has a need or desire to use a game for learning the material, or whether the game will likely meet a high level of resistance. This user group information can be useful in determining various aspects of the game design from the themes of the art and the style of gameplay to the complexity of language.

Small, manageable items such as difficulty of language or controls can potentially have a large impact on user behavior. Use of an inappropriate level of language can quickly alienate the user and negatively affect the user's desire to use the program as well as their effectiveness at understanding and using the tool. Controls, too, can enhance or diminish the experience. Poor implementation of either of these can negatively impact the desired learning outcome. Overwhelming the user with difficult language, confusing game structure for their skill level, and excessive information can greatly tax the resources of the user, and will likely result in more effort being exerted in figuring out the definitions and controls and less resources available to focus on learning objectives.

2.2 Story Boarding and Paper Prototyping

During this phase of development, designers should begin developing concepts for the game in storyboard format, including game design, style, and art while taking general target population and SME considerations into account. Basic principles of human factors and user-centered design should be adhered to while constructing the proposed game model. Generating a chart that lists objectives, game features, implementation, and outcomes can be helpful in this process.

Researchers during this phase of development should create a paper prototype of the game as developers design it. A paper prototype is essentially a paper-based version of the game that is capable of demonstrating the structure and elements of the game. This prototype should follow the storyboards and should progress just as the game would from screen to screen. Paper prototyping can be a very useful and cost-effective tool for assessing user reaction to wording, layout, and sequencing or flow of game progression [8], and can be used to experiment with different ideas. Paper prototypes require little technical experience to utilize and are quick and easy to create and modify, unlike computer prototypes that can take a great deal of time and technical skill to modify [9]. Creating the prototype also allows for the first exposure of the game concept to the target users. You should conduct a small focus group of five or six that includes SMEs and members of the target population who play the paper prototype. The information gathered from the prototype trials can provide useful insight to the modifications that should be made prior to software development, and can allow for changes that make the game more useful and accessible to the user. It is important to collect general information and opinions about the game as it stands, the structure, characters, and overall presentation, as well as basic usability feedback that can be helpful in correcting simple errors before further development.

During paper prototyping questions should focus on game features, general usability, and ease of understanding. Game features include style of art and gameplay as well as the narrative and plot. Usability issues at this time are largely conceptual and less specific: whether the control scheme seems to fit the game and makes sense, whether the screen order and progression seem logical, and whether the objectives seem clear. Many of these questions can be culled from numerous sources that we have found useful, including items from the System Usability Scale (SUS) [10], Questionnaire for User Interface Satisfaction (QUIS) [11], and Technology Acceptance Model [12]. The items relate to perceived usefulness, behavioral intention, ease of use, application-specific self-efficacy, enjoyment, opinion of game elements, general usability and playability, and player preferences.

At this stage questions can either be asked directly with varying amounts of detail, or a Likert-scale rating system can be employed. To use a Likert scale, we ask that the tester rate each of these items on a 5-point Likert scale, where a 1 indicates "strongly disagree" and a 5 indicates "strongly agree." Scores on Likert scales can be averaged individually or for each subscale.

Focused follow-up questions to further understand the testers' responses should be asked after the questionnaire to collect further information about their problems or suggestions. It is important to take feedback into consideration when there are specific problems that the testers encounter. It is also important, however, to avoid trying to satisfy each individual's design suggestions, as their personal preferences may not represent the general population and this will only add more work to the development process than necessary.

2.3 Build Alpha 1

Once you have updated the storyboards with the feedback from the focus groups, the programmers create a rough first alpha build of the game. This version of the game lacks any sort of finished art or sound design and serves as an extension of the paper prototype. It should be complete in function only.

2.4 Usability for Alpha 1

Upon completion of a working computer draft, an in-house "game breaking" session should take place. This should be a one-day session involving the researchers and other individuals involved with incorporating usability design principles in the game. Include SMEs in this process if the target audience may have problems operating the game, comprehending the material, or if they pose any additional special needs that may be difficult for the researchers to identify on their own. The "game breaking" process simply involves testing the functionality and limitations of the game. By using members within the organization, testing is relatively fast and affordable compared to recruiting groups of target users and individually directing them through formal testing. In-house sessions do not necessarily call for any form of administered paper measure, but the evaluators should keep questions from the measures in mind, as well as basic human factors and design principles, as they assess the usability of the game. This procedure allows the development team to detect a number of potential problems that may be missed during playtesting due to limited time and the sometimes inhibited exploration that can come with closely observed behavior. It can also provide additional insight into the problems that testers may encounter, allowing for deeper understanding of tester responses during user trials. Errors and bugs found should be logged into a bug tracker, and issues that are determined to be violations of human factors and usability principles, such as difficult controls and fuzzy text, should also be recorded for the programmers to review.

After the in-house testing has been completed, a small usability study with as many as five participants pulled from an easily accessible population, such as college students, should be conducted. These participants should all be of good health and sound physical condition. Since this is not conducted in the target population, the goal of this round of usability testing is to catch the glaring usability issues so that future testing can focus on the more succinct, population-specific issues [13]. This phase of testing should take approximately three days to allow time to run participants and collect data. Any required paperwork for participation, such as informed consent, should be completed by testers before interacting with the game. A survey to collect background information should also be administered. Upon completion of this survey, a brief assessment of topic-specific knowledge should be conducted to provide pretest scores. Each participant should then be independently exposed to the game in order to observe opinions that are uninfluenced and unbiased by other participants.

There are several methods of administering usability testing, each with their own unique goals. There are those that involve goals to be completed within the game to determine if users can complete a given task in the game with instructions as well as those that involve free exploration of the software to examine how intuitive and easy to use the program is without any guided instruction. While interacting with the game, an observer records their behavior on paper, in particular when participants are having difficulties, asking questions, or demonstrating strong emotion.

Using a think-aloud protocol can also be informative to understanding exactly how the user is approaching the task. It can help identify where their focus is drawn and what tasks they find difficult to carry out. Developed by Ericsson and Simon [14], it involves the tester speaking aloud their thoughts and actions as they interact with the game. This procedure can provide useful internal information as to how the user feels or thinks in real time that may not be captured by post-test paper measures. Video recording test sessions can also be very informative and thorough, allowing for excellent future reference, but this can also prove very time-consuming to analyze especially when the sessions are long [8, 15]. We do not recommend this practice as a session as brief as a half hour can result in over six hours worth of transcription and analysis [16]. While there are advantages to think-aloud studies, there are also several disadvantages, including interference with play due to increased cognitive load, potentially difficult to interpret and unquantifiable data, as well as the threat of altering the flow of the experience. This detrimental phenomenon can be especially apparent when trying to assess engagement and enjoyment, as the talking task disrupts the game-playing experience [16]. Due to the positive and negative aspects of each testing method, it may be beneficial to administer a combination of goal-focused testing with the think-aloud protocol. A modified think-aloud study provides task-specific goals that testers must achieve in addition to instructing participants to speak their thoughts only when they are having difficulty, if they need to ask for instruction, or if they encounter something worth commenting upon. This can minimize interference of the think-aloud process while still allowing insight into the thought process of the user.

Learning objectives, like playability, may be difficult to assess this early on in the development process. If there are numerous bugs and usability issues, this will greatly detract from the learning outcome by increasing the effort to operate the game, thereby decreasing the available effort and attention focused on learning the material. Again, this should be kept in mind when conducting the research, and when determining whether to assess learning goals at this stage. If you do decide to assess learning, the learning assessment should be conducted after playing the game. If teaching declarative knowledge is the goal, multiple choice and matching questions are adequate for assessing learning outcomes. If, however, the learning goals are intended to be dynamic or metacognitive in nature, the assessment format should be free response questions or completion of related transfer tasks. Collecting this data will allow for pretest/post-test observations of changes in performance and can help determine whether the learning goals are being achieved during each stage of testing. If there is strong reason to believe that learning goals will be seriously affected by conducting simultaneous usability measures, such as a large amount of content or highly complex content covered, additional participants can be recruited who only play the game and complete the learning assessment.

Once the learning assessment is completed, surveys should be distributed to the participants for completion. The length and order of questionnaires distributed can also prove to be important when conducting usability research. Very long surveys can diminish patience in participants, resulting in a lack of willingness to complete surveys—especially after long, aggravating sessions of usability trials [10, 11]. When participants have been performing tasks for long periods of time, their responses to questionnaires can become influenced by their fatigue or desire to be done with testing, which may result in inaccurate data being collected and items being skipped. To avoid some of these problems, it is important to try to make the questionnaires as relatively short as possible while still receiving the necessary feedback.

When conducting usability testing, we use sections from both the SUS and the QUIS that we have updated and modified for the gaming domain, as well as additional items from various other scales. The SUS was generated to provide quick assessment of usability without taxing the tester for long periods of time [10]. Though it provides a set of validated measures concerning usability, it was not designed with serious games in mind, and thus it lacks greater depth that is desired when conducting gaming research. The QUIS on the other hand is a very extensive usability scale that consists of many specific areas of usability and numerous questions. It has been scaled back in length on several occasions, but once again focuses largely on systems usability rather than gaming usability. For that reason, additional enjoyment and engagement questionnaires are combined in order to collect data on playability. Depending on the amount of bugs and usability issues expected, it may be difficult to collect meaningful data about the playability at this stage. Engagement and flow, two subscales associated with playability, can be greatly hampered by issues of usability and the presence of bugs. Furthermore, the unfinished quality of the game, art, and sound, for example, may also elicit negative responses. These scales are therefore not of great importance this early on, though they may still provide insight and give a starting score to compare with later builds of the game.

Once all the necessary data has been collected a one-day process of data entry and analysis should take place. Bugs found by testers should once again be recorded in the bug tracker. Qualitative and quantitative data should be analyzed and included in the report for the programmers, and should outline findings with interpretations, usability issues, and specific suggestions for improvements.

2.5 Build Alpha 2

The programmers need to be provided with the usability report. The researchers' suggestions may not always be practical and working with developers directly can help the team decide upon more simple yet effective solutions. Programmers will then be allotted time to address and fix all of the issues found during testing in the Alpha 1 build. This version still features place-holding artwork, yet is functionally complete, usable, and free of most major bugs.

2.6 Usability for Alpha 2

A second round of usability testing will be conducted once again in small groups of about five individuals, however this time the individuals will be pulled from the target audience. This is to ensure that any population-specific usability issues are addressed. At this stage, learning objectives will also be reassessed to ensure that the game is not only usable, but effective. The usability testing will proceed in the same manner as Alpha 1, with only slight differences. These differences are mostly in the additional learning assessment and playability scales that are implemented if they were not incorporated earlier. In terms of interpretation, these scales are now far more important during this stage of development. For these reasons, testing from the representative population is desirable, as these scales have more meaning when applied to the desired testing group. Additional focus questions may also be included in the interview portion where it is determined that more user feedback is desired. Again, a report is prepared and presented to programmers.

2.7 Build Beta

After all major issues from the report have been corrected, development enters the Beta development stage. During this stage, the majority of intense development occurs. For example, all of the art is completed and placed in the game, all menus are finished, and all elements of the whole game experience are included and complete.

2.8 Usability for Beta

This round of usability testing will once again draw upon approximately five individuals from the target population to ensure that all issues have been addressed. The process is identical to that of Alpha 2 usability testing. A final report is prepared with any lingering issues and is provided to the programmers.

2.9 Final Build

Any issues still present from the beta usability study are addressed and a final draft of the game is prepared. Then conduct one final round of usability testing that will take place solely in-house using researchers and on-staff usability experts. This is to ensure that the final draft is free of any and all bugs, that all issues have been fixed, and that all of the goals of the game have been met. Any last minute adjustments should be implemented; however, at this stage, the game is ultimately complete.

3 Conclusion

Usability testing is a critical process necessary for developing effective serious games. With the addition of measures for playability and learning outcomes, it is possible to improve the design process while also ensuring the development of a successful training and learning tool. Our procedure provides structure and measures that are both effective and efficient in the development of well-rounded serious games. The procedure also accounts for the restrictions of small serious game development companies specifically, providing a process that is time-efficient and cost-effective, while requiring minimal personnel.

We hope to provide the industry of serious games a useful design process to improve the quality and success of their product. By incorporating effective development principles and usability testing, serious games can ensure the development of games that are both functional and successful in producing the desired learning outcomes.

References

1. Crawford, C.: Lessons from Computer Game Design. In: Laurel, B. (ed.) The Art of Human-Computer Interface Design, pp. 103–111. Addison-Wesley, Reading (1999)
2. Resnick, H., Sherer, M.: Computerized Games in the Human Services – An Introduction. In: Resnick, H. (ed.) Electronic Tools for Social Work Practice and Education, pp. 5–16. The Haworth Press, Bington (1994)
3. Witmer, B.G., Singer, M.J.: Measuring Presence in Virtual Environments: A Presence Questionnaire. In: Presence: Teleoperators and Virtual Environments, vol. 7, pp. 225–240. MIT Press, Boston (1998)
4. Jackson, S.A., Eklund, R.C.: The Flow Scales Manual. Fitness Information Technology (2004)
5. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The Development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. J. Exp. Soc. Psychol. 45, 624–634 (2009)
6. Dillman, D.A.: Mail and telephone surveys: The total design method. Wiley, New York (1978)
7. Sudman, S., Bradburn, N.B.: Asking Questions: A practical guide to questionnaire design. Jossey-Bass Publishers, San Francisco (1982)
8. Shneiderman, B., Plaisant, C.: Evaluating Interface Designs. In: Shneiderman, B., Plaisant, C. (eds.) Designing the User Interface, pp. 140–171. Pearson Education, Boston (2005)
9. Medero, S.: Paper Prototyping. A List Apart (January 23, 2007)
10. Brooke, J.: SUS: A Quick and Dirty Usability Scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London (1996)
11. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument measuring user satisfaction of the human-computer interface. In: SIGCHI 1988, New York, pp. 213–218 (1988)
12. Yi, M.Y., Hwang, Y.: Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and technology acceptance model. Int. J. Hum.-Comput. Stud. 59, 431–449 (2003)
13. Krug, S.: Don't Make Me think: A Common-Sense Approach to Web Usability. New Riders Publishing, Indianapolis (2006)
14. Ericsson, K.A., Simon, H.A.: Protocol Analysis: Verbal reports as data. Bradford Books/MIT Press, Cambridge (1984)
15. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond human-computer interaction. John Wiley & Sons, New York (2002)
16. Hoonhout, H.C.M.: Let the game tester do the talking: Think aloud and interviewing to learn about the game experience. In: Isbister, K., Schaffer, N. (eds.) Game Usability: Advice from the Experts for Advancing the Player Experience, pp. 65–77. Morgan Kaufmann, New York (2009)

Experience-Based Curiosity Model: Curiosity Extracting Model Regarding Individual Experiences of Urban Spaces

Chihiro Sato, Shigeyuki Takeuchi, and Naohito Okude

Keio University Graduate School of Media Design,
4-1-1 Hiyoshi, Kohoku, Yokohama, Japan
{chihiro,take,okude}@kmd.keio.ac.jp
http://www.kmd.keio.ac.jp

Abstract. Many online advertising and web-based recommendation systems have been developed, however not so many services consider individual's real activities in the real world for real time recommendation, regarding the experience of particular person. Environmental sensing from mobile devices has become capable of understanding the environment by sensing from mobile devices; though they do not necessary interact with the people directly. We present Experience-based Curiosity Model, a model indicating individual's real time curiosity within the city regarding how well the individual knows the city. It aims to understand individual's real time interests by not relying on information the people input intentionally but by understanding behavior data. This paper evaluates the model with this sensor device prototype, and elaborates possibilities when understanding individuals in detail by extracting the curiosity predicted from current behaviors using sensors.

Keywords: Urban Experience, User Analysis, Curiosity, Urban Experience, Ethnography, Behavior.

1 Introduction

In this paper, we present Experience-based Curiosity Model, a model indicating individual's real time curiosity within the city regarding how familiar the individual is with the city. It aims to understand individual's real time interests by not relying on information the people input intentionally but by understanding behavior data. It calculates individual's real time curiosity level by analyzing users' behavior depending on users walking speed, collected from sensors, as TTI Model [17], a model extracting individual's curiosity level in urban spaces on their spare time by collecting behavior data from sensors when spending time within the city has executed. Experience-based Curiosity Model uses utility theory and Bayesian Networks [16] to understand the experience the individual has already had in the past with the certain city, and automatically switches the model's calculation method of the curiosity depending on the familiarity.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 635–644, 2011.

As many online advertising and recommendation systems have been developed for web-based marketing, customers utilize these systems to search for their needs and actually buy whatever that catches their hearts. These internet based recommendation web sites enable individuals to simplify the strokes of shopping around to find anything that interests them, for information have clearly been overloaded everywhere [19]. On the other hand, the concept of participatory sensing [3], environmental sensing from mobile devices has leveraged personal devices to gather data from everyday mobile devices such as cellphones in domains of personal experience sharing, public health, environmental monitoring, or anything else to form interactive sensor networks enabling public and professional users to collect, analyze and share local data. Researches such as networked sensor system measuring environmental factors such as air pollution or noise to be used by students, parents, bicyclists and the homeless [10], and constructing both hardware and software for gathering raw data by attaching sensors to GPS-enabled cell phones in order to understand how urban air pollution impacts both individuals and communities have been done [6].

Not so many of these streams, however,consider individual's real activities in the real world for real time recommendation, regarding the experience of the particular person. There have been new applications using individual's current behavior information directly [15], but not considering to understand individual's real activities in the real world have been constructed for real time recommendation. The prior participatory sensing works are capable of understanding the environment by sensing from mobile devices, however they do not necessary interact with the people directly.

Experience-based Curiosity Model will be able to calculate both types of users, depending on how often the individual comes to the certain place, by changing the probability table of the transition of the actions, for it was motivated by the fact found from an ethnographic research in Tokyo Station, the largest terminal station in Tokyo. We prototyped a sensor device to be worn on the users' waists consisting of 3D accelerometer sensor module and GPS sensor. This paper evaluates the model with this sensor device prototype, and elaborates possibilities when understanding individuals in detail by extracting the curiosity predicted from current behaviors using sensors.

2 Related Studies

2.1 Mobile City Guide Applications

Magitti [1] generates recommendations of content matching from user activities without having to issue queries targeting young urbanites. CityFlocks [2] facilitates social navigation in urban public spaces both directly and indirectly utilizing average ratings and comments from local residents by categorizing and retrieving location in urban places, by using tags generated by users to show comments or recommendations. CitySense [12] enables people in San Francisco to find out hot-spots and where people-like-me are now in the city by providing a map based summary of current hot-spots summarized with clustering the analyzed real time feeds of activity using the collected GPS data in smartphones to understand everybody's daily activity and detect anomaly outbroken or elevated clusters. These previous works enable users in urban space to find out what place may be suitable for the user by categorizing or filtering the user's activities on purpose, to be used for urban navigation.

2.2 Recommendations by Activity-Detecting

Krumm presents a novel method for predicting the location of a driver's destination during the drive [9]. The prediction based on the common intuition which drivers tend to chose efficient routes, is shown as a probability along with a map of driving times to compute the probability of any candidate destination from a database of driving trips they gathered with GPS receivers. Drivers use the prediction as a guide to decide which information to automatically present to the driver, depending on where the driver is going. Schechtner *et al.* developed a recommendation system aiming to influence people's movement patterns within a national park for conserving [18]. User's real time location data is collected using GPS sensors, and route suggestions or other information designed to influence the user's movement in the park is provided to the user on a PDA screen. These works predict user's behaviors and use the data for suggesting what routes or ways to go next moment.

2.3 Context Awareness Modeling

CenceMe shares inferences of the presence of individual information sensed from the mobile phones through social networking applications [13,14]. The presence of the user's is captured as a status in terms of the activity, disposition, habits, and surroundings. It is an application that combines the inference of the presence of individuals using off-the-shelf, sensor-enabled mobile phones, and works on a system called AnonySense [5,7]. AnonySense is architecture for applications based on collaborative, opportunistic sensing with privacy by using personal mobile devices. AnonySense allows applications to submit sensing tasks that will be distributed across anonymous participating mobile devices. This receives the verified sensor data reported back from the field, and provides the first secure implementation of an participatory sensing model. There is an underlying threat model and trust model of AnonySense, and the location-blurring feature provides statistical k-anonymity.

3 Concept

3.1 Model Extracting Individual's Curiosity Level

TTI (Time Transient Interests') Model [17], a model extracting individual's curiosity level in urban spaces on their spare time by collecting behavior data from sensors, calculates individual's real time curiosity level by analyzing the walking speed within the city. The model was produced from a case study focusing on a female university student taking a random walk to spend her spare time in Tokyo, with a 3D accelerometer and a GPS sensor taped on her arm also connected to a laptop held by another person. From this case study, a hypothesis that "walking slowly when finding anything interesting and stopping after the gradual decrease of the speed, the curiosity level may be quite high" has been formulated. To strengthen this hypothesis, an observation in Tokyo Station, the largest terminal station in Tokyo has been pursued. When taking a close look at the people there spending their spare time, they walked heading towards the next store, which they lose speed and stop again. From this observation, the hypothesis was strengthened, and also added that when walking at a rather fast constant speed, a person does not have much interest at the moment.

Fig. 1. Corsage Device Prototype

TTI model runs within a corsage-like sensor device to be worn on the users' waist, consisting 3D accelerometer and GPS sensors to capture the speed change of an individual, synthesized with Arduino Pro mini [1] with java. The synthesized data is converted by FFT at a speed of 64 set of data per second to derive the power spectrum of accelerometer as other research executes [4]. The derived value of power spectrum will be categorized every second by Support Vector Machine (SVM) algorithm into four groups; standing still, walking slowly, normal speed, and walking fast.

Coordinating with Bayesian Networks, consisting of a graphical structure and a probabilistic description of the relationships among the variables in a system [8], TTI model is used to calculate the probability of interest when sensor data is collected from the person every second. When inputting the data, the model which applies Russell's filtering [16] extracts the noise of sensors through the link under a probability of normal distribution, connected with the node of the action; standing still, walking slowly, normal speed, or walking fast. The interest probability considers the change of speed, regarding the action taken the second before.

To verify the model, two field evaluations were executed with one particular female university studentfor 20 minutes. The interview was implemented to the female student by asking her to evaluate the places which she actually felt interested with value between 0 and 1. The feedback that we obtained from the

[1] Arduino http://www.arduino.cc

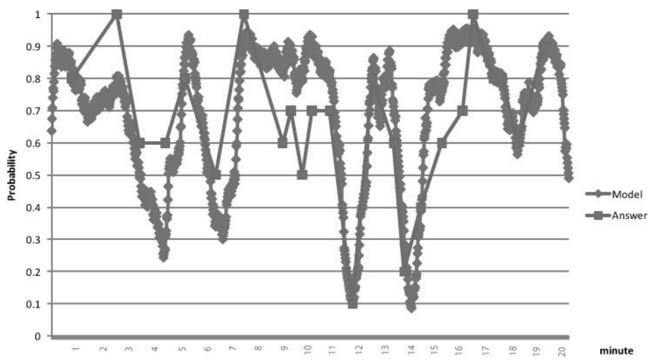


Fig. 2. TTI Prediction and Interview Answers

interview answers corresponded with the probabilities TTI Model had calculated, as can be seen in the comparison graph (See Figure 2) of her curiosity predicted from the model.

In order to implement out that TTI Model works with not just only one particular person, a field evaluation was executed with 20 university students to spend their spare time for an hour at Ebisu, a popular town for young adults in Tokyo. We had every student wear the corsage device on their waist, attached to a laptop, and recorded every person's random walks with a video camera. We also wrote down thick descriptions about the characteristics of the individual's activity patterns, and drew where the students walked and seemed to have interest in directly on a paper map. The students were asked to walk around the town as if there were alone.

Figure 3 is the result of the interest probability tagged with the GPS sensor to the map of Ebisu, all combined together. The areas that have a high tower is an area that people felt interest in, according to the TTI Model.

Fig. 3. 20 Students' Curiosities in the city

3.2 Ethnographic Research of TTI Model Evaluation

As mentioned earlier, we executed interviews based on Lazar's method of research methods [11] with each and every users, which showed relevance as the preceding two tests did. We conducted prior interviews asking four questions 1) How often do you come to this town? 2) At what times do you take random walks? 3) What kinds of things do you feel interest? and 4) What kind of image do you have for this town? We also conducted interviews after the hour of random walk, asking questions 1) What are the places you still have in mind from the walk? 2) Did you find anything that you want to tell someone else? 3) Did you find anything that you will want to come again? The places that we received from the posterior question 'What are the places you still have in mind from the walk' corresponded with the results the model outputted.

We realized that the students who had answered the interview as they frequently come to this town, their average walking speed seemed to be faster than those who answered they were unfamiliar to this town. The *new-comers* of this town walk around taking a look at almost anything that come into their come into their visions, though the *experts* of the town know the town very well, so they do not pay so much attention to each and everything in the space. However, when these *experts* actually stop their feet to take a look at something, their curiosity level is quite high. From this evaluation test, we formulated a hypothesis of "the activities of experts and non-experts of the city changes" and "the difference between experts and non-experts can be determined by how often they come to the area" was formulated to refine the TTI Model.

To support this hypothesis, we have pursued an ethnographical observation in Tokyo Station once again. Observing the department store within the station where many people spend their spare time waiting for the next train, we found there were two types of people; people walking through the department store as a passageway and people wandering around and window-shopping. The people wandering around are all walking slowly, and slows down when finding anything that caught their attention. When taking a closer look, they started walking again heading towards the next store, which they lose speed and stop again. However, the people who use the station on a daily basis walk on a faster average speed, and when those who are experts of the place actually lose speed and stop, they really did find something that catch their hearts.

3.3 Experience-Based Curiosity Model

Experience-based Curiosity Model will be able to calculate both types of users, depending on how often the individual comes to the certain place, by changing the probability table of the transition of the actions.

In addition to TTI Model, Experience-based Curiosity Model includes utility theory as decision networks for understanding how interested the individual is and their satisfactions. The interest level decision node is influenced by how frequent the individual comes to the certain city, and also how crowded the city is at the moment. This is because of the observations which we realized that

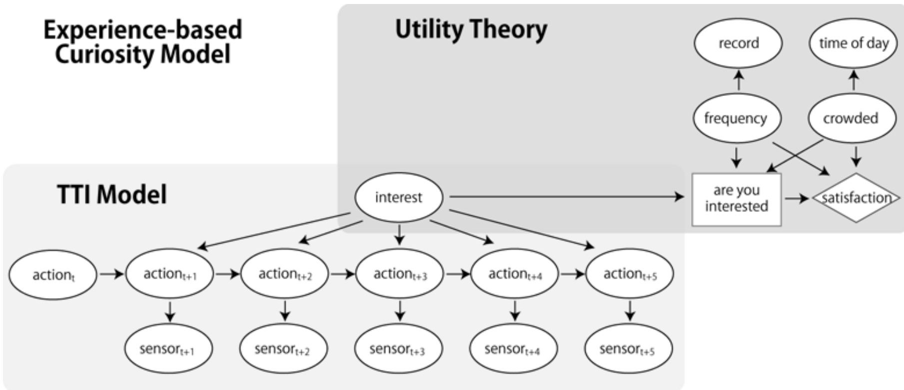


Fig. 4. Experience-based Curiosity Model

the activities of *experts* and *non-experts* differ. We have also considered the fact that if the city is too crowded, as most cities in Tokyo are at the rush hours, the individual is forced to walk slower than usual due to traffic jams. The frequency could be understood by analyzing records of the individual, and the crowdedness can be sorted according to the time of day, for example 6 PM is rush hour for most cities in Japan therefore the city is quite likely to be crowded. The records are implemented by the interviews that we executed beforehand about how often the individual comes to the city, at the moment.

Experience-based Curiosity Model also runs within a corsage-like sensor device to be worn on the users' waist, consisting Arduino Pro mini synthesized 3D accelerometer sensor module "KXM52-1050" and GPS sensors with java. Undergoing FFT and SVM, the derived value of power spectrum will be categorized every second nto four groups; standing still, walking slowly, normal speed, and walking fast. Utilizing *Netica* [2], a commercially available java based modeling software, Experience-based Curiosity Model coordinates Bayesian Networks and utility theory as Figure 4 shows.

4 Field Evaluation

Utilizing the raw data of the field evaluation we had executed in Ebisu with 20 students, we had managed to calculate the curiosity of individuals accurately than TTI's calculations, according to the interviews and analysis of video recorded activities. Figure 5 shows one example of an *expert* of this town, for Experience-based Model (EC) dictates a rather lower probabilities in general, and extracts certain places that the person actually stopped their feet to take a look.

The map on the left of Figure 6 is the curiosity map of one particular *expert* of the city according to the Experience-based Curiosity Model, and the one of

[2] Norsys Software Corp. http://www.norsys.com

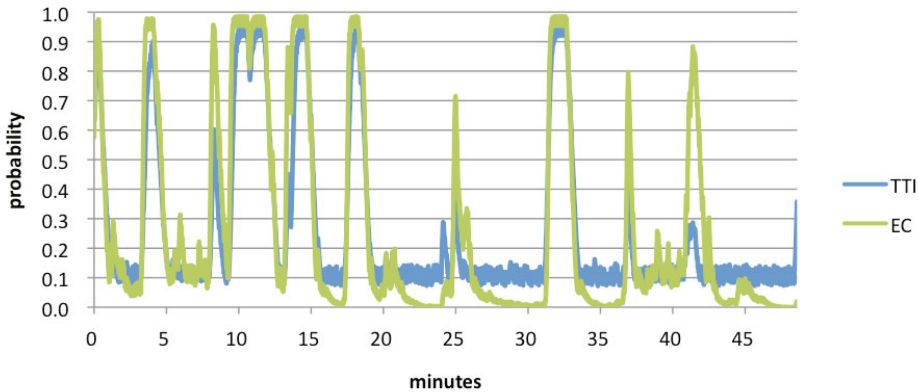


Fig. 5. Comparison Graph of the Experience-based Curiosity Model

the right is implemented by TTI model. Higher probabilities are shown in red, while the lower probabilities are shown as blue. We have succeeded in keeping down the curiosity level of the experts of the city, whereas they usually do not keep attention throughout their visit in their *home town*. The places shown as red in the left map corresponds with the interview results of the individual.

Fig. 6. Comparison Map of the Experience-based Curiosity Model

5 Conclusion and Discussion

Experience-based Curiosity Model will be able to calculate both types of users, depending on how often the individual comes to the certain place, by changing the probability table of the transition of the actions, for it was motivated by the fact found from an ethnographic research in Tokyo Station. We have prototyped a sensor device to be worn on the users' waists consisting of 3D accelerometer sensor module and GPS sensor. We evaluated the model with this sensor device prototype, extracting curiosities of individuals according to their experiences in the city and the current behaviors. As future works, we would like to use the *SUICA*[3] activity records for the input of record for understanding the familiarity

[3] http://www.jreast.co.jp/suica

of the city, for in Japan, nearly every person has an IC card to ride trains or buses. It stocks 20 frequent records of what station they rode on and off the train. As an average of the 20 students' SUICA records we took, the 20 records holds about one month of the records therefore privacy matters would not matter so much.

Experience-based Curiosity Model will enable navigation applications in the city to become more efficient depending on the how familiar the individual is with the city. New-comers of the city will be able to understand how the experts of the city spend their time, and expand their experience although it will be their first visit. Experience-based Curiosity Model will enhance experience of people's urban lives.

Acknowledgments. We thank any anonymous reviewers who gave us helpful comments.

References

1. Bellotti, V., Begole, B., Chi, E.H., Ducheneaut, N., Fang, J., Isaacs, E., King, T., Newman, M.W., Partridge, K., Price, B., Rasmussen, P., Roberts, M., Schiano, D.J., Walendowski, A.: Activity-based serendipitous recommendations with the magitti mobile leisure guide. In: CHI 2008: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp. 1157–1166. ACM, New York (2008)
2. Bilandzic, M., Foth, M., De Luca, A.: Cityflocks: designing social navigation for urban mobile information systems. In: DIS 2008: Proceedings of the 7th ACM Conference on Designing Interactive Systems, pp. 174–183. ACM, New York (2008)
3. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.: Participatory sensing. In: World Sensor Web Workshop, pp. 1–5. Citeseer (2006)
4. Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P.P., Koscher, K., LaMarca, A., Landay, J.A., LeGrand, L., Lester, J., Rahimi, A., Rea, A., Wyatt, D.: The mobile sensing platform: An embedded activity recognition system. IEEE Pervasive Computing 7(2), 32–41 (2008), 1399120
5. Cornelius, C., Kapadia, A., Kotz, D., Peebles, D., Shin, M., Triandopoulos, N.: Anonysense: Privacy-aware people-centric sensing. In: Proceeding of the 6th International Conference on Mobile Systems, Applications, and Services, pp. 211–224 (2008)
6. Honicky, R., Brewer, E.A., Paulos, E., White, R.: N-smarts: networked suite of mobile atmospheric real-time sensors. In: NSDR 2008: Proceedings of the Second ACM SIGCOMM Workshop on Networked Systems for Developing Regions, pp. 25–30. ACM, New York (2008)
7. Kapadia, A., Triandopoulos, N., Cornelius, C., Peebles, D., Kotz, D.: AnonySense: Opportunistic and Privacy-Preserving Context Collection. In: Indulska, J., Patterson, D.J., Rodden, T., Ott, M. (eds.) PERVASIVE 2008. LNCS, vol. 5013, pp. 280–297. Springer, Heidelberg (2008)
8. Kjærulff, U.B., Madsen, A.L.: Probabilistic networks - an introduction to bayesian networks and influence diagrams. Aalborg University and HUGIN Expert A/S (2005)

9. Krumm, J.: Real time destination prediction based on efficient routes. SAE 2006 Transactions Journal of Passenger Cars - Electronic and Electrical Systems (2006)
10. Kuznetsov, S., Paulos, E.: Participatory sensing in public spaces: activating urban surfaces with sensor probes. In: DIS 2010: Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 21–30. ACM, New York (2010)
11. Lazar, J., Feng, J., Hochheiser, H.: Research methods in human-computer interaction. Wiley, Chichester (2009)
12. Loecher, M., Jebara, T.: Citysense: multiscale space time clustering of gps points and trajectories. In: Joint Statistical Meeting (2009)
13. Miluzzo, E., Lane, N.D., Eisenman, S.B., Campbell, A.T.: CenceMe – injecting sensing presence into social networking applications. In: Kortuem, G., Finney, J., Lea, R., Sundramoorthy, V. (eds.) EuroSSC 2007. LNCS, vol. 4793, pp. 1–28. Springer, Heidelberg (2007)
14. Miluzzo, E., Lane, N.D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S.B., Zheng, X., Campbell, A.T.: Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application. In: Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, pp. 337–350 (2008)
15. Nakajima, T., Lehdonvirta, V., Tokunaga, E., Kimura, H.: Reflecting human behavior to motivate desirable lifestyle. In: DIS 2008: Proceedings of the 7th ACM Conference on Designing Interactive Systems, pp. 405–414. ACM, New York (2008)
16. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Inc., Englewood Cliffs (2009)
17. Sato, C., Takeuchi, S., Imbe, T., Ishibashi, S., İnami, M., Inakage, M., Okude, N.: Tti model: model extracting individual's curiosity level in urban spaces. In: DIS 2010: Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 352–355. ACM, New York (2010)
18. Schechtner, K., Schrom-Feiertag, H.: Understanding and influencing spatio-temporal visitor movement in national parks based on static and dynamic sensor data. In: Pervasive 2008 Workshop Proceedings, pp. 95–99 (2008)
19. Wei, Y., Moreau, L., Jennings, N.: A market-based approach to recommender systems. ACM Transactions on Information Systems (TOIS) 23(3), 227–266 (2005)

Implied Aesthetics: A Sensor-Based Approach towards Mobile Interfaces

Daniel Sauter

Program Coordinator, New Media Arts
School of Art and Design, University of Illinois at Chicago
dsauter7@uic.edu

Abstract. This article focuses on the implied aesthetics resulting from the affordances of 4th generation mobile devices, and in particular: motion-awareness. It draws connections between web standards, open source platforms, and their implications on diverse output media. Motion-awareness is rooted in the possibility to detect and interpret device orientation and acceleration — thus by extension the movements and gestures of individuals interacting with handheld devices. The article introduces the *KETAI* platform, designed to aid mobile applications that rely on motion analysis. Faster and more accurate detection of device attitude, context, and transportation mode enables a variety of novel applications in the traffic, gaming, and heath care sectors.

Keywords: Mobile interface, user experience, motion-awareness, sensor platform, gyroscope.

1 Introduction

Ubiquitous computing represents, after mainframe computing and personal computing, the "third paradigm"[1] where computation is integrated into the built environment. Souza e Silva denotes cell phones as "interfaces of hybrid spaces" [1], asserting a merger between physical and imaginary spaces through nomadic technologies. This article focuses on the implied aesthetics resulting from the affordances of 4th generation mobile devices now entering the product mainstream, and in particular: motion-awareness. A smaller sibling to geolocation, motion-awareness is rooted in the possibility to detect and interpret device orientation and acceleration — thus, by extension, the movements and gestures of individuals interacting with handheld devices. Electronic sensors that pertain to motion-awareness include accelerometer[2], magnetometer[3], and gyroscope[4] sensors. Fused with additional sensors, including "proximity sensor, accelerometer, touch input panel, ambient light sensor, ambient noise sensor, temperature sensor, gyroscope, a hinge detector, a position determination device, an orientation determination device, a motion sensor, a sound sensor, a

[1] (Weiser, 1996), referencing Allan Kay's term for the third era of computing.
[2] Sensing orientation and acceleration relative to g-force, including vibration, shock, or falling.
[3] Measures orientation in relation to the earth's magnetic field.
[4] Measures rotation around axis, able to detect rotation around gravity (in contrast to accelerometer).

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 645–654, 2011.

radio frequency electromagnetic wave sensor, and other types of sensors and combinations thereof" [2], handheld devices are destined to evolve into motion-aware game consoles and universal remote controls. Particularly mobile augmented reality and navigation applications can benefit from data fusion between accelerometer and gyroscope, allowing detection of device orientation and rotation around gravity. Jumping, chipping, riffing, and drumming — Nintendo Wii has already conditioned millions of end-users with the digital haptics of gesture-based human computer interaction (HCI) during family pastime. Introduced in 2006 with built-in three-axis accelerometer and infrared motion tracking, the addition of the MotionPlus [3] dual-axis gyroscope in 2008 ultimately completed "the missing piece of the puzzle" [4]. Gyroscope-equipped[5] 4th generation cell phones, tablet computers and personal digital assistants are now challenging this application space. Motion-awareness leverages gesture-based interaction principles towards a portable, public, and ubiquitous space, with applications reaching from traffic, to gaming and health care. The *KETAI*[6] software platform [5] introduced in this article is designed to aid the development of mobile applications and services that rely on native sensors and motion analysis.

2 Context

The Xerox Palo Alto Research Center's Computer Science Laboratory coined the term ubiquitous computing in 1988, proclaiming: "Just as electric motors have disappeared into the background of everyday life, PARC scientists have envisioned a future mobile computational devices will be similarly transparent" [!6]. Mark Weiser, chief scientist at Xerox PARC described ubiquitous computing as "roughly the opposite of virtual reality". In this model, computers "live out here in the world with people," as opposed to "putting people inside computer-generated worlds". Weiser's highest ideal was the "invisible" interface: "so imbedded, so fitting, so natural, that we use it without thinking about it" [7]. The intimate language suggests a very personal device — disappearing like a piece of undergarment. Weiser did not assert gradual societal adaptations as a catalyst for technological 'transparency' or 'invisibility', as HCI expert Donald Norman did in *The Invisible Computer* (1998). Bolter and Gromala refuted the "myth" of transparency in 2003 as a "story that artists and designers have told us (and themselves) in order to justify their designs" [8]. In contrast, PARC's definition #1 clearly constitutes ubiquitous computing as "different from PDA's, dynabooks, or information at your fingertips. It is invisible, everywhere computing that does not live on a personal device of any sort, but is in the woodwork everywhere" [7]. Weiser's deliberate differentiation between "everywhere computing" on the one hand, and "information at your fingertips" through personal devices on the other, is challenged in this article. The majority of the motion-aware applications described here do not adhere to this dichotomy. They are designed to move seamlessly between embedded infrastructure, "the cloud", and mobile devices.[7] The "inch-, foot-, and

[5] The gyroscope was introduced to Apple iOS4 on June 24, 2010, Samsung Galaxy S Android followed July 15, 2010.

[6] *Ke-tai* is the term for cell phone used in Japan, it means: "extension of one's hand".

[7] "Seamlessness" references "calm technologies" (Weiser's and Seely Brown's,1995), referring to computational devices that move seamlessly from the periphery to the center of people's attention.

yard-sized computers" coined "Tabs, Pads, and Boards"[8] [7] in the early 1990s affirm how influential PARC's vision still is today.

3 Implied Aesthetics

This article draws connections between web standards and their implications on diverse output media. It proposes an implied aesthetics that arises in conjunction with the sensory affordances of 4th generation mobile devices. Due to the scope of this article, no attempt will be made to trace individual histories. Rather, particular aspects of HTML5, CSS3, JavaScript, jQuery [10], Processing.js [11], and mobile sensors are juxtaposed to outline their combined innovative potential, resulting in HTML5-based, JavaScript-processed, and sensor-powered mobile applications. The notion of implied aesthetics is considered 'impartial' in this context. It is not reacting to a visual style, but acknowledges a momentary sense of innovation, followed by the successive transformation into a visual norms and tropes soon after.

Hereby, the inch-, foot-, yard-sized computers consumed today pose particular design challenges, often rooted in the range of fragmented standards, device sizes, and screen resolutions. These client-side 'unknowns' can either be accommodated by common-denominator compromises, or, multiple media- and device-specific solutions[9]. Fragmentation through multiple solutions has never been a desirable prospect, as it requires redundant efforts for the same problem, hence additional costs and room for error. Standardized and customized solutions are held in balanced by the number of targeted output media and the afforded native implementation detail. Centralized databases have propelled another 'unknown' through the separation of design and content, emphasizing durable content alongside adaptable design. Hence, the database as "symbolic form"[10] has played a crucial role in the predominance of dynamic web sites, web applications, and native apps that scale seamlessly in size, performance, resolution, storage, bandwidth, and platform. When groups of co-creators 'come together' to collaboratively author documents, centralized version control systems (VCS [9]) play an important role to facilitate cooperation, administrate hierarchies, and negotiate common goals and standards.

Because all major web browsers support the HTML5, CSS3, and JavaScript standards, web applications represent a competitive choice for developing cross-platforms applications. On the other hand, native applications provide access to hardware-specific features, as for instance device sensors and camera(s). To bridge this gap, application frameworks such as Appcelerator Titanium and PhoneGap support native development[11] via JavaScript and HTML5, with the goal to leverage versatility and productivity towards native features.

---

[8] i.e. *Tablet Computers* (Microsoft Corp., 2000), *iPad* (Apple Inc., 2010), and *Galaxy Tab* (Samsung, 2010).

[9] Paralleled by the dialectic between open and proprietary industry standards such as XHTML/CSS and Macromedia/Adobe Flash.

[10] Referencing Lev Manovich's writings on the database as symbolic, cultural form, and genre of new media (1998-2001).

[11] Tinanium compiles HTML5-JavaScript code into native apps, PhoneGap makes native feature available to create HTML5-JavaScript-based web apps.

Technically, the abovementioned design tools do not constitute an implied aesthetics per se. These standards deploy access to standardized HTML as well as native UI elements — hence maintain a hybrid visual canon of 'extended' standards. While CSS3 and JavaScript provide comprehensive tools to manipulate visual properties, the Canvas[12] element represents the most versatile, and for this line of argument most influential addition to the HTML5 package. The element is designed to render scriptable drawings and images, and therefore the least predefined, most customizable, and aesthetically neutral. In the discussion of the most influential visual design additions, one should also mention the CSS3 @font-face property, providing the first comprehensive cross-browser solution for embedding non-system fonts into web sites and applications. Both elements eliminate a significant legacy of compromises for designers and developers, opening up animation features that were previously implemented within proprietary containers. In 2006, John Resig released the first version of jQuery, "a fast and concise JavaScript Library that simplifies HTML document traversing, event handling, animating, and Ajax interactions for rapid web development" [10]. Now used by Google, Dell, Netflix, Mozilla, etc., it has grown into a web standard in its own right. In 2008, Resig ported the Processing[13] language to JavaScript using the Canvas element [11], opening up development to "tens of thousands of students, artists, designers, researchers, and hobbyists who use Processing for learning, prototyping, and production" [12]. A selection of HTML5 creations and applications is collected at Chrome Experiments under the tag line: "Not your mother's JavaScript" [13]. Tens of thousands of artists, designers, and researchers utilizing Processing will arguably play an important role in making the Canvas element, and the web browser, a site for visual innovation [14-16]. The resulting aesthetics are implied by the following aspects: a) the possibility to render complex data visualizations and generative designs via JavaScript, combined with b) the possibility to retrieve data asynchronously from the server in the background (Ajax). Powered by significantly improved JavaScript performance (particularly via WebKit and V8 engines), HTML5 and JavaScript-based web applications are well equipped to challenge the status quo of visual culture on the Web. This is particularly the case for applications that can leverage GPU-based graphics processing, including GPU-accelerated WebGL 3D computer graphics.

GPU-accelerated 2D and 3D real-time graphics are destined to produce a significant visual impact on (web) applications in both exiting, and exuberant ways (see "rich" animations of the 1990). Co-creation and open source principles have produced about 100 Processing extensions [17] over the last decade, extending into open hardware, data protocols, sound, and simulations. This number is a testament to the collaborative power of online communities, and a foretaste of what can be adapted to native handheld applications.

The 'native' user experience is largely defined by the refined degree of haptics and interactivity due to by native hardware. Augmented reality (AR) applications that leverage pitch, roll and yaw angles (gyroscope) in conjunction to g-force (accelerometer) is one example where decades of research can bow be leveraged due to ubiquitous mobile devices. Pointing the device camera towards the horizon and following

[12] Introduced by the WebKit open source web browser engine.
[13] Developed by Ben Fry and Casey Reas in 2001.

the horizon line is a typical scenario where the gyroscope can contribute to detect rotation around gravity, instead of relying on the magnetometer's attempt to find a reference point (geographic north). Wikitude [21] and Spot Crime [22] augment a 'world browser' and crime locations onto a live image, one example out of many mapping and navigation use cases that could be improved by more accurate device attitude. Sun Seeker [18], DishPointer [19], and See Breeze [20] augment the sun, satellite location and wind. A detailed discussion of educational [23] and gaming scenarios [24], as well as human factors (finger tips: 9.2 mm for discrete and 9.6 mm for serial tasks [25]) exceeds the scope of this article.

4 KETAI Platform

Ubiquitous mobile devices equipped with gyroscope and accelerometer sensors enable a variety of novel applications. Targeted for sectors such as traffic [26-27], gaming [24], health care [28,29], and public safety, *KETAI MOTION* is a software framework designed for approximating the physical device path. This approximation is based on a dynamic data fusion method that captures and analyzes data from accelerometer, magnetometer, and gyroscope sensors, complemented by sensors that are relevant within a particular context (e.g. light sensors and digital cameras). Determining vehicle speed, sudden breaking, or collecting statistical data on traffic congestion [35] requires a flexible software framework that accommodates dynamic data fusion for multiple sources. Deployed on 4th generation mobile devices, it is designed to capture, abstract, archive, and analyze device attitude, movement, and gestures. *KETAI* manages data streams, and makes those resources available to other applications as a content provider. It is designed to extract motion features and differentiate between modes of transportation, gait, kinematic velocity, and hand gestures. It can dynamically reduce data granularity based on power or performance settings. By fusing gyroscope, accelerometer, and magnetic field sensors, *KETAI* complements existing interpretation methods of existing research, predominantly relying on geolocation [30-31, 53]. The ability to distinguish between device rotation and translation is the foundation for "Wii-like" [32] mobile gaming applications. The *KETAI* platform offers a novel approach to motion analysis, building on increased data granularity. The platform allows to emphasize resources and accuracy for the most relevant sensors in a given device context. For instance, if a mobile device is carried in a pocket or bag, only a low level of data granularity is required (reserving resources), whereas if a user is looking at the screen (face recognition), it enables a higher level of granularity. In comparison to existing work in the context of motion-awareness, *KETAI* offers the following advantages:

Integration. The platform is capable to register and manage all electronic sensors (e.g. accelerometer, gyroscope, magnetic field, light, temperature, etc.). As a platform, it remains adaptable for a variety of applications. It offers dynamic data fusion and provides machine learning (Hidden Markov Model (HMM)) and signal processing algorithms (i.e. onset detection). A content provider for other services on a device, it makes interpretative data available to other applications.

Abstraction. The platform analyzes device attitude vectors in real-time, extracting points of interest based on detected changes via motion analyzer (Fig. 1). By dynamically removing data granularity, and adding time-stamped metadata, the framework can continuously capture and summarize device movement at the desired level of accuracy. The ability to shift sensor and power resources opportunistically towards particular sensors allows *KETAI* to limit capture, analysis, and resources.

Aggregation. The platform will enable motion comparison between multiple devices in close proximity via unregulated wireless networks [33-35], enabling assertions about transportation mode etc., by enumerating devices and comparing their relative motion difference. If proximate devices share similar device attitude vectors, the transportation mode might output “car”, “bus”, or “train”. If a significant amount of devices share nearly identical attitude vectors, the transportation mode can be refined to “bus” or “train”. In the gaming context, this feature supports applications that rely on real-time interaction and collaboration.

Interpretation. By integrating (front- and back-facing) digital cameras into the *KETAI* platform, assertions can be made about device context and gestures. Frame differencing between front- and back-facing camera images affords a distinction between device rotation and hand gestures. Face recognition provides an additional contextual resource for interpreting device context.

Gyroscope and accelerometer data is continuously refined by magnetometer and geolocation sources. Existing work related to device motion [36, 37, 38, 39, 29, 40], transportation mode [41, 28, 42], gait [43-49], or device context [50-52], does not meet our application goals. These systems often rely on external devices [43, 44],

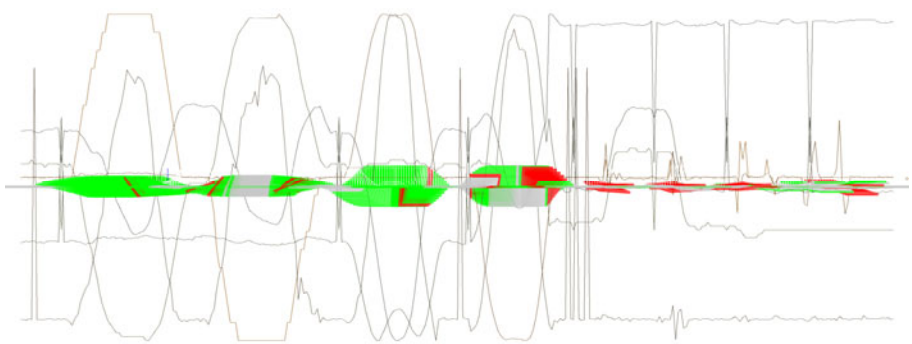


Fig. 1. Sensor fusion visualized by *KETAI MOTION* on Android OS[14] over the duration of 60 seconds (horizontal axis). Figure is based on accelerometer and gyroscope data, showing device orientation and signal peaks.

[14] While not limited to a specific operating system, a preliminary version of the *KETAI* library for Processing has been implemented for the Android OS, first released at the Processing.Android Conference (http://processingandroid.org) on Oct. 2nd, 2010, at the University of Illinois at Chicago Innovation Center (organized by the author). Android OS accounts for 53 percent (Feb. 2011 [58]) of smartphones sold in the U.S.

hence are inconvenient and not ubiquitous. They offer sufficiently accurate data only for a limited range of use cases, and lack accuracy due to a single [30-31, 53], or two combined sensor sources [36, 41]. Preexisting classification models transportation mode rely predominantly on geolocation [30-31, 53], or accelerometer combined with GPS data [36]. The addition of the gyroscope adds the crucial level of improved accuracy to determine transportation mode and device context. Furthermore, detection of rotation around gravity [21] allows novel applications in the context of gesture recognition [54] and gait analysis [43-46].

While processing real-time sensor data, *KETAI MOTION* determines "points of interest", motion peaks, duration, and momentum [55]. Augmented with timestamps (millisecond accuracy), device attitude vectors describe the first layer of abstracted data. Abstraction allows managing, packaging and compression of motion data over extended periods of time. The information is then stored locally in SQLite [56] databases, enabling post-analysis, and intermittent data transfer to the cloud.

5 Conclusion

Web standardization and comprehensive support of HTML5 (Canvas), CSS3 (@font-face), and JavaScript (jQuery), represents the basis for visually innovative, interactive (Ajax), dynamic (Processing.js), and rich (WebGL), high-performance (GPU-accelerated) applications. The large developer base already utilizing those standards will play a major role in the transition towards a new software-based visual culture. The most innovative approaches successfully merge native device features such as gyroscope sensors. With the onset of multi-touch displays as ubiquitous interfaces the mouse button and rollovers became obsolete. Motion sensors can play a role in complementing some of the absent, mouse-related 'pseudo classes'. Within the operating system for example, leveling and tapping the device can assist in scrolling applications where the screen would otherwise be obstruct by fingertips. Motion- and context-aware applications can seamlessly toggle perspectives for mobile navigation applications. Transportation mode detection can improve safety and provide transportation-relevant information for a particular geolocation.

The *KETAI* platform leverages the affordances of 4th generation mobile devices, and in particular motion-awareness. It manages all built-in electronic sensors, focusing on gyroscope, accelerometer, and magnetic field sensor, offering a comprehensive approach for detecting motion and gesture. Faster and more accurate information about device attitude, context, and transportation mode enables a variety of novel use cases, ranging from traffic, to gaming, and heath care. Seventy percent of executives state that their companies regularly create value through Web communities [59, 60]. Open source platforms such as Processing(.js), Titanium, WebKit, and Android, have never played a more pivotal role in leading the way towards a mobile visual culture.

References

1. Souza e Silva, A.: From cyber to hybrid: mobile technologies as interfaces of hybrid spaces. Space & Culture 9(3), 261–278 (2006)
2. Apple Inc., United States Patent 7,633,076 (1999)

3. Nintendo's Wii MotionPlus Controller, http://www.nintendo.com/wii/console/accessories/wiimotionplus
4. InvenSense IDG-600/650 dual-axis gyroscope, http://invensense.com/mems/gaming.html
5. Sauter, D.: KETAI Sensor Platform, http://ketaimotion.com
6. Weiser, M.: http://www.ubiq.com/hypertext/weiser/quicktime/UbiCompIntro.qt
7. Ibid, http://sandbox.xerox.com/ubicomp
8. Bolter, D., Gromala, D.: Windows and Mirrors: Interaction Design, Digital Art, And the Myth of Transparency. The MIT Press, Cambridge (2003)
9. Apache Subversion Subversion open source version control system, http://subversion.apache.org/
10. JQuery, The Write Less, Do More, JavaScript Library, http://jquery.com/
11. Resig, J.: Processing.js, http://ejohn.org/blog/processingjs/
12. Fry, B., Reas, C.: Processing.org, http://processing.org
13. Chrome Experiments, http://www.chromeexperiments.com/
14. Chrome Experiments — WebGL Experiments, http://www.chromeexperiments.com/webgl/?f=webgl
15. Inside Toronto Article: Seneca student project has real-world implications, http://www.insidetoronto.com/community/education/article/931731-seneca-student-project-has-real-world-implications
16. Exhibition Archives, Processing, http://processing.org/exhibition/
17. Processing.org Libraries, http://processing.org/reference/libraries/
18. Sun Seeker: 3D Augmented Reality Viewer for iPhone 3GS, iPhone 4, and iPad on the iTunes App Store, http://itunes.apple.com/us/app/sun-seeker-3d-augmented-reality/id330247123?mt=8
19. DishPointer Augmented Reality for iPhone and iPod touch (4th generation) on the iTunes App Store, http://itunes.apple.com/us/app/dishpointer-augmented-reality/id323135933?mt=8
20. See Breeze: 3D Augmented Reality Wind Visualizer for iPhone 3GS, iPhone 4, and iPad on the iTunes App Store, http://itunes.apple.com/us/app/see-breeze-3d-augmented-reality/id366765248?mt=8
21. WIKITUDE for iPhone 3G, iPhone 3GS, and iPhone 4 on the iTunes App Store, http://itunes.apple.com/us/app/wikitude/id329731243?mt=8
22. Spot Crime for iPhone, iPod touch and iPad on the iTunes App Store, http://itunes.apple.com/us/app/spotcrime/id347343610?mt=8
23. Inkling — Interactive Textbooks for iPads, http://www.inkling.com/
24. Eliminate: GunRange for iPhone, iPod touch and iPad on the iTunes App Store, http://itunes.apple.com/us/app/eliminate-gunrange/id377754042?mt=8
25. Parhi, P., Karlson, A., Bederson, B.: Target size study for one-handed thumb use on small touchscreen devices. In: MobileHCI 2006, pp. 203–210 (2006)
26. BlueToadTM, TrafficCast, http://trafficcast.com/products/bluetoad/
27. Op.cit., http://trafficcast.com/docs/BlueTOAD_Overview_October_2010.pdf
28. Nyan, M., Tay, F., Murugasu, E.: A wearable system for pre-impact fall detection. Journal of Biomechanics, 3475–3481 (2002)

29. Li, Q., Stankovic, H.M., Barth, A., Lach, J.: Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information. In: BSN 2009, pp. 138–143 (2009)
30. Patterson, D., Liao, L., Fox, D., Kautz, H.: Inferring High-Level Behavior from Low-Level Sensors. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 73–89. Springer, Heidelberg (2003)
31. Liao, L., Patterson, D., Fox, D., Kautz, H.: Learning and inferring transportation routines. Artif. Intell. 171, 311–331 (2007)
32. Nintendo: Wii Console, Wii Motion Plus, `http://www.nintendo.com/wii`
33. Wolfson, O., Xu, B., Yin, H.: Dissemination of spatial-temporal information in mobile networks with hotspots. In: Ng, W.S., Ooi, B.-C., Ouksel, A.M., Sartori, C. (eds.) DBISP2P 2004. LNCS, vol. 3367, pp. 185–199. Springer, Heidelberg (2005)
34. Sistla, P., Wolfson, O., Xu, B.: Opportunistic Data Dissemination in Mobile Peer-to-Peer Networks. In: Anshelevich, E., Egenhofer, M.J., Hwang, J. (eds.) SSTD 2005. LNCS, vol. 3633, pp. 346–363. Springer, Heidelberg (2005)
35. Xu, B., Wolfson, O., Naiman, C., Rishe, N.D., Tanner, R.M.: A Feasibility Study on Disseminating Spatio-temporal Information via Vehicular Ad-hoc Networks. In: Proc. of the 3rd International Workshop on Vehicle-to-Vehicle Communications, V2VCOM (2007)
36. Agrawal, S., Constandache, I., Gaonkar, S., Choudhury, R.: PhonePoint Pen: Using Mobile Phones to Write in Air. In: MobiHeld 2009 (2009)
37. Sakaguchi, T., Kanamori, T., Katayose, H., Sato, K., Inokuchi, S.: Human motion capture by integrating gyroscopes and accelerometers. In: IEEE/SICE/RSJ, pp. 470–475 (1996)
38. Kunze, K., Lukowicz, P.: Dealing with sensor displacement in motion-based onbody activity recognition systems. In: UbiComp 2008, pp. 20–29 (2008)
39. Li, Q., Young, M., Naing, V., Donelan, J.: Walking speed estimation using a shank-mounted inertial measurement unit. Journal of Biomechanics, 1640–1643 (2002)
40. Lorincz, K., Chen, B., Challen, G. W., Chowdhury, A. R., Patel, S., Bonato, P., Welsh, M.: Mercury: A Wearable Sensor Network Platform for High-Fidelity Motion Analysis. In: SenSys 2009, pp. 1–14 (2009)
41. Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Using Mobile Phones to Determine Transportation Modes. ACM Transactions on Sensor Networks 6(2), article 13 (2010)
42. Reddy, S., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Determining Transportation Mode On Mobile Phones. In: 12th IEEE International Symposium on Wearable Computers, pp. 25–28 (2008)
43. Mariani, B., Hoskovec, C., Rochat, S., Bula, C., Penders, J., Aminian, K.: 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. Journal of Biomechanics, 2999–3006 (2010)
44. Aminian, K., Najafia, B., Bula, C., Leyvraz, P.-F., Robert, P.: Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. Journal of Biomechanics, 689–699 (2002)
45. Nyan, M.N., Tay, F.E.H., Seah, K.H.W., Sitoh, Y.Y.: Classification of gait patterns in the time–frequency domain. Journal of Biomechanics, 2647–2656 (2006)
46. Saitou, K., Horikawa, E., Saitoh, H., Masuda, T.: Tri-axial Measurements of Gait with Miniature Gyroscope Sensors. Journal of Biomechanics, 535 (2007)
47. Mayagoitia, R., Nene, A., Veltink, P.: Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems. Journal of Biomechanics, 537–542 (2002)

48. Center for Embedded Networked Sensing: Walking Speed Estimation with Gaussian Process Regression on Inertial Sensors. Annual Report, Participatory Sensing (PART), 113–142 (2010)
49. Vathsangam, H., Emken, A., Spruijt-Metz, D., Sukhatme, G.: Toward Free-Living Walking Speed Estimation Using Gaussian Process-based Regression with On-Body Accelerometers and Gyroscopes. In: PervasiveHealth 2010, pp. 1–8 (2010)
50. Wissen, B., van Palmer, N., Kemp, R., Kielmann, T., Bal, H.: ContextDroid: an Expression-Based Context Framework for Android. In: PhoneSense 2010, 6–10 (2010)
51. Lin, C.-Y., Chen, L.-J., Chen, Y.-Y., Lee, W.-C.: A Comfort Measuring System for Public Transportation Systems Using Participatory Phone Sensing. In: SenSys 2009, pp. 31–35 (2009)
52. Miluzzoy, E., Papandreax, M., Laney, N.D., Luy, H., Campbell, A.T.: Pocket, Bag, Hand, etc. — Automatically Detecting Phone Context through Discovery. In: SenSys 2009, pp. 26–30 (2009)
53. Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.: Understanding mobility based on GPS data. In Ubiquitous Computing, pp. 312–321. ACM, New York (2008)
54. Wilson, A., Shafer, S.: Demonstration of the XWand Interface for Intelligent Spaces. In: UIST 2002 Companion, pp. 37–38 (2002)
55. Dixon, S.: Onset Detection Revisited. In: 9th Int. Conference on Digital Audio Effects, pp. 1–6 (2006)
56. SQLite self-contained, serverless, zero-configuration, transactional SQL database engine, http://www.sqlite.org/
57. Open Source Computer Vision library of programming functions for real time computer vision, http://opencv.willowgarage.com/wiki/Welcome
58. Fortune Tech: Android's U.S. market share hits 53%, http://tech.fortune.cnn.com/2011/01/31/npd-android-os-phones-now-outsell-everyone-else-in-us-combined/
59. How companies are benefiting from Web 2.0: McKinsey Global Survey Results (September 2009), http://www.mckinseyquarterly.com
60. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch (August 2010), http://www.mckinseyquarterly.com

A Study on the Expected Image and Relevant Design Techniques in Different Product-Use Stages

Yung-Chin Tsao, Brian Chen, and Yen-Pang Yang

Graduate School of Design Science, Tatung University,
40 Zhongshan North Road, 3rd Section, Taipei 104, Taiwan
bchen1969@gmail.com

Abstract. This study is an exploratory investigation on user's expectation emotions and design techniques associated with product use. A behavior unitization method has been introduced for dividing product-use behavior into 3 stages: starting, development, and ending. The sub-concepts of expected image in different stages were found by the study. The design elements of each sub-concept has been further analyzed, and the corresponding design techniques were generated through the focus group discussion. There are total 18 combinations of design techniques gained from this study.

Keywords: Expected image, Behavior unitization, Product use process.

1 Introduction

Today, the competition in consumer market is intense and the emotional factors of product usage have become more and more important for customers. Therefore, companies start to consider the emotional responses of users and implement those responses into new product development. It is expected that designs with the considerations of product usage will satisfy the consumer because of the pleasure feelings obtained by the customers.

Expectation refers to the inner thought and eager expectation of individual and makes one toward positively and aggressively. When someone had the so-called expecting feeling to something, it will make people feel even more eager about the thing he expected. In other words, expectation can cause positive results when those expecting image has been implemented into product design.

About the expectation, Wilson has proposed "The Affective Expectation Model" and he described the expectation is"how people predict their feelings in specific scenarios or coming specific stimulation."[1] In emotional three-dimensional model proposed by Plutchik(1980), expectation was one of 8 basic emotions of human beings. Besides, expectation included the levels of intensity. The expectation with lower intensity was the curiosity about something or was interested in things. Maximum intensity of expectation meant someone extremely cared about the results of the things and was on the alert status.[2] The one of purposes of this study is to explore the expected image construct in different product-use stages.

In terms of product-use stage, Edward Reed has proposed a method for dividing people's everyday actions into units, and found the everyday actions can be described as a hierarchical organization. [3] Taso and Hsiung explored the relationship between

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 655–663, 2011.

the adverbial images of complex actions in using utensils and the forms of the utensils. It is found that the behavior of use is organized by a series of single actions. And different process of use may result in different image. [4] It is important that how to divide a behavior into units, Chou has suggested 3 stages for describing a story: starting, development, and ending. [5] Figure 1 illustrates an example for the behavior unitization: the behavior "Ming goes to school in the morning" can be divided into 3 stages. The starting stage is "Ming gets up in the morning". The development stage is "Washing face, changing clothes and being ready to school". The ending stage is "Going to the school". Each stage can be divided into more detailed actions. However, in this study the 3 stages will be the target for dividing people's behavior.

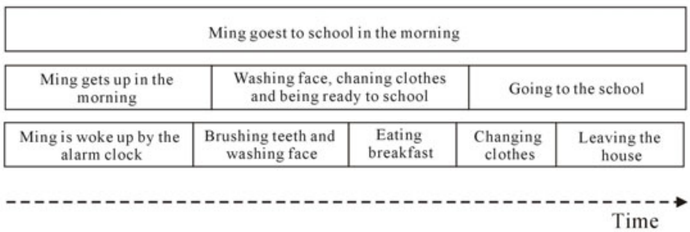


Fig. 1. A behavior unitization example

2 Investigating the Expected Image Constructs in Different Stages

2.1 Case Examples of Expected Image in Different Stages

An open-ended questionnaire is used for the investigation of case examples.

There were a total of 29 subjects (age 20 30; 13 males and 16 females) participated the questionnaire investigation. The subjects were asked to describe the expected incidents or scenarios in the past living experience and the subjects were also instructed to divide the whole process into three stages (starting, development, and ending) according to the sequence of time. In total, 49 case examples were collected from the questionnaire. After that, a focus group discussion was held to discuss and screen the results of questionnaires. 31 expected cases were gained after screening out of 49 cases from the questionnaires for further evaluation.

2.2 Conceptual Structure of the Expected Image in Different Stages

A similarity degree evaluation was used to further understand the expected image constructs. The evaluators were 7 students with design background. The results acquired were categorized by cluster analysis which further led to three different conceptual structures of the expected image (Figures 2 to 4). Descriptions of sub-concept in different stages are shown in Tables 1.

The starting stage of expected image was constructed by two sub-concepts: "attractive incentives" and "expectation on the results". The keywords of "attractive incentives" were "selection", "discovery", "temptation", and "fresh". The keywords of "expectation on the results" were "purposeful", "joyful", "impatient", "imagination", and "fantasy".

The development stage of expected image was constructed by three sub-concepts: "cautious", "impatient" and "excited". The keywords of "cautious" were "depressed", "uneasy", "worried", "attempt", "cautious" and "sequential". The keywords of "impatient" were "tight" and "nervous". The keywords of "excited" were "expectation" and "excited".

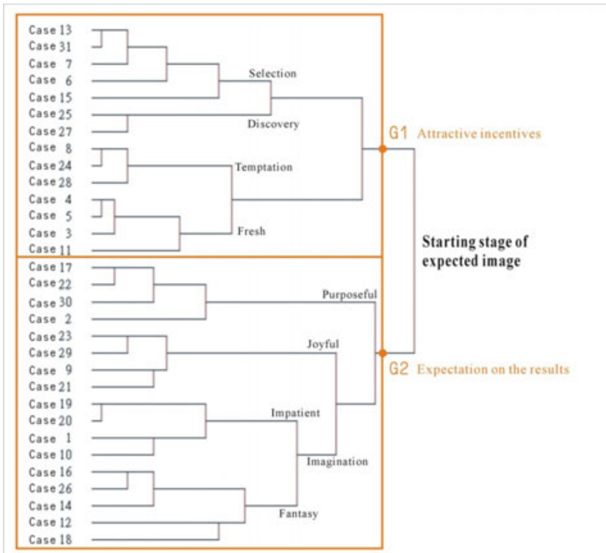


Fig. 2. Conceptual structure of the starting stage of the expected image

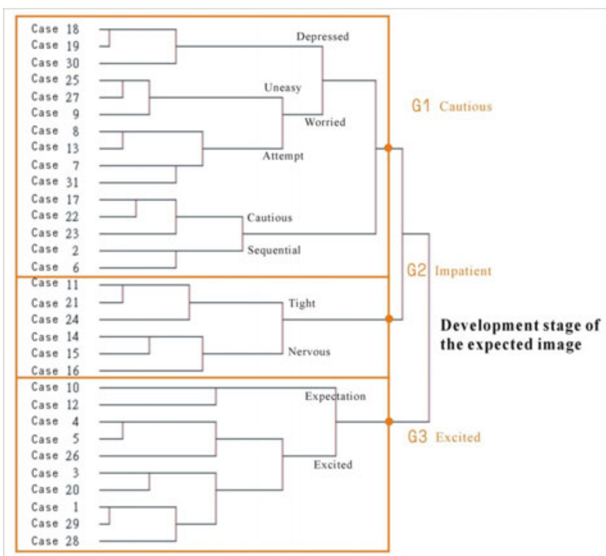


Fig. 3. Conceptual structure of development stage of the expected image

The ending stage of expected image was constructed by three sub-concepts: "different from the expected", "going to be revealed" and "meeting the expectation". The keywords of "different from the expected" were "illusion", "dissatisfied", "disliked", and "disappointed". The keywords of "going to be revealed" were "immediate" and "known". The keywords of "meeting the expectation" were "happy" and "satisfied".

It is found that the 3 stages of expected image are constructed by sub-concepts. However, the cases used for the constructs are daily-live cases. The next step is to investigate product-use cases for each sub-concept in different stages.

Table 1. Sub-concept and descriptions of expected image in different product-use stages

	Sub-concept	Descriptions
Starting stage of expected image	G1:Attractive incentives	Being tempted by fresh things and it is the charm which cannot be resisted.
	G2: Expectation on the results	There is a specific goal for accomplishment and imagination with the results.
Development stage of expected image	G1: Cautious	Cautious and careful process with uneasy state.
	G2: Impatient	Being excited and impatient with the results
	G3: Excited	The state in high spirits.
Ending stage of expected image	G1: Different from the expected	Finding it is different from the expected
	G2: Going to be revealed	Almost certain state before the result. (from the view of emotional levels)
	G3:Meeting the expectation	The result showed is as expected and people feel satisfied and delighted

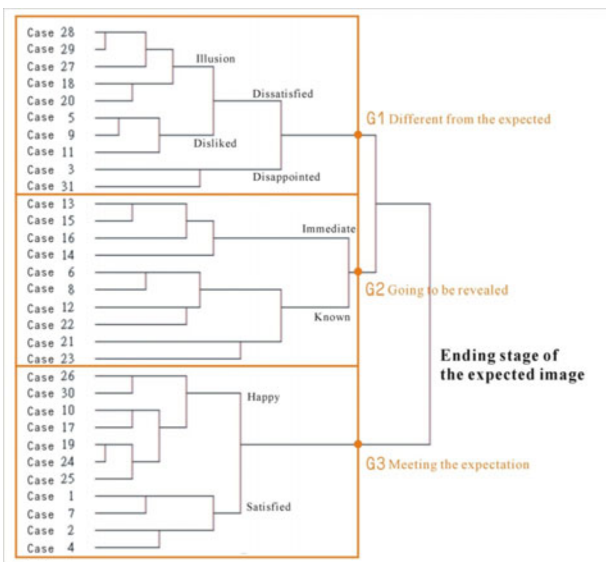


Fig. 4. Conceptual structure of ending stage of the expected image

3 Design Technique Investigation and Analysis in Different Product-Use Stages

3.1 Product-Use Case Collection of Sub-concept of the Expected Image in Different Stages

An open-ended questionnaire was used to collect the product-use cases base on the sub-concepts of the expected image. 19 subjects (age 20~30, 10 males and 9 females, all the subjects have design experiences) were asked to describe the product cases which meet the items of expectation sub-concepts in different stages, and reasons were required to answer. 19 questionnaires were completed and after screening, there were 15 to 21 cases included in each sub-concept. These cases will be used for the evaluation of important design elements.

3.2 Important Design Elements in Different Product-Use Stages

According to Tsao and Chen's study on "The Expected Image in the Process on Using Products"[6], it is found that there were four design elements would affect the expectation in product using process. These four design elements are "exterior form, material presentation, function, and scenario for using process". The design element "scenario for using process" has already been explored and divided into 3 stages for further study in this research. In order to get more detailed information, the design element "function" has been separated into "function return" and "operational method".

Base on the above setting, four design elements "exterior form, material presentation, function return and operational method" were used to conduct a 10 scale evaluation on each related product-use case collected from the questionnaire (1 meant no influence and 10 meant the most influential). The result is shown as Table 2.

It is found that in the starting stage for using products, "exterior form" and "material presentation" were more important than "function return" and "operational method". This means product appearance plays an important role at the beginning of product-use process to bring the emotion of expectation. In development stage, "function return" and "operational method" were ranking higher than "exterior form" and "material presentation". It shows the function and operation is more important than appearance of a product, to keep the expectation emotion during the development stage. However, in ending stage, "function return" and "exterior form" were found to be more important than "material presentation" and "operational method". It indicates the result of product form and feedback of function to be the key-points of the expectation emotion for the ending stage.

3.3 Design Techniques in Different Product-Use Stages

A focus group discussion was held to discuss and analysis the ranking result of design element importance. The purpose of discussion was to generate corresponding design techniques for each expectation sub-concept, as shown in Table 3.

Table 2. The ranking of importance of sub-concept of expectation in 3 product-use stages

Stages	Sub-concept of expectation	Ranking	Design elements	Importance	Descriptions of case characteristics	Generalized design techniques
Starting stage of expected image	Attractive incentives	1	Exterior form	7.29	Unique exteriors form	The exterior form is different from the past characteristics
		2	Material presentation	6.47	The materials different from the products of the same type	
		3	Function return	5.23	Combining new functions and services	
		4	Operational method	5.05	New and diverse use methods	
	Expectation on the results	1	Exterior form	7.14	Exterior form which is easy to identify with and understand	The exterior form implies the purposes
		2	Material presentation	5.57	Natural materials	
		4	Function return	5.28	The following reaction can be expected	
		3	Operational method	5.35	The operation is easy to understand	
Development stage of expected image	Cautious	3	Exterior form	3.80	Appearance implies danger	Careful operational process
		4	Material presentation	3.71	The material characteristic tends to lead to negative results	
		1	Function return	7.14	Messages leading to failure or danger	
		2	Operational method	7.04	Careful and sequential operation	
	Impatient	3	Exterior form	3.94	Increasing or sequential model	The implication of suspended results
		4	Material presentation	3.27	Presentation supporting appearance or functions	
		2	Function return	6.72	Counting down characteristic	
		1	Operational method	7.11	Suspended	
	Excited	3	Exterior form	4.50	Interesting associations	Back-and-forth operational method
		4	Material presentation	2.68	Leading to returns	
		1	Function return	7.62	Interesting or tight returns	
		2	Operational method	7.06	Back-and- forth or exaggerated	
Ending stage of the expected image	Different from the expected	2	Exterior form	7.22	Symbolizing certain functions	Presenting the things without image
		4	Material presentation	3.88	The same as outer conditions	
		1	Function return	7.77	Different from the messages delivered by the appearance	
		3	Operational method	4.72	Natural operation	
	Going to be revealed	2	Exterior form	7.25	Implicative and it must be understood by thinking	Ambiguous result presentations
		4	Material presentation	3.25	Different from the past	
		1	Function return	7.66	Indirect	
		3	Operational method	6.66	Repetitive actions	
	Meeting the expectation	2	Exterior form	7.35	Clear and definite presentation	strengthening result presentations
		4	Material presentation	2.92	Meeting outer conditions	
		1	Function return	7.57	Direct and natural	
		3	Operational method	4.71	Proceeding with instructions	

Table 3. Design techniques corresponding to the 3 stages

Stages	Expectation sub-concepts	Corresponding design techniques	Descriptions of design techniques
Starting stage for using products	Attractive incentives	The exterior form is different from the past characterstics	We cannot predict the functions or use purposes of the products simply from appearance.
	Expectation on the results	The exterior form implies the purposes	The appearance is smiliar to present products so that people can predict the functions and operational methods.
Development stage for using products	Cautious	Careful operational process	The users must finish specific conditions step by step to use the products and they should be careful in each step.
	Impatient	The implication of suspended results	After having simple operation, the users should wait for a certain period of time until the result is revealed.
	Excited	Back-and-forth operational method	When the users operate the products, the products will have immediate feedback.
Ending stage for using products	Different from the expected	Presenting the things without image	The final result of the product is unrelated to the previous two stages in terms of meanings.
	Going to be revealed	Ambiguous result presentations	It is indirect, encoded, and it must be absorbed through thinking.
	Meeting the expectation	strengthening result presentations	The final result of the product is related to the previous two stages in terms of meanings.

The corresponding design technique for "attractive incentives" is "the exterior form is different from the past characteristics", it means people cannot predict the function or use purpose of the product simply from it's appearance. The corresponding design technique for "expectation on the results" is "the exterior form implies the purposes", it means the appearance is similar to present product, so that people can predict the function and operational method. The corresponding design technique for "cautious" is "careful operational process", it means people must finish specific conditions step by step to use the product and should be careful in each step. The corresponding design technique for "impatient" is "the implication of suspended results", it means after having simple operation, people should wait for a certain period of time until the result is revealed. The corresponding design technique for "excited" is "back-and forth operational method", it means when people operate the product, the product will have immediate feedback. The corresponding design technique for "different from the expected" is "presenting the things without image", it means the final result of the product is unrelated to the previous two stages in terms of meanings. The corresponding design technique for "going to be revealed" is "ambiguous result presentations", it means the result is indirect, encoded, and it must be absorbed through thinking. The corresponding design technique for "meeting the expectation" is "strengthening result presentations", it means the final result of the product is related to the previous two stages in terms of meanings.

3.4 Combination of Design Techniques in Different Stages

According to the findings, there are total 8 design techniques corresponding to each expectation sub-concept in the 3 product-use stages; 2 for the starting stage, 3 for the development stage, and 3 for the ending stage. Thus it can become a combination of design techniques base on time sequence, as shown in Figure 5. There are total 18 possible combinations of design techniques which are relevant to expected image of product-use process.

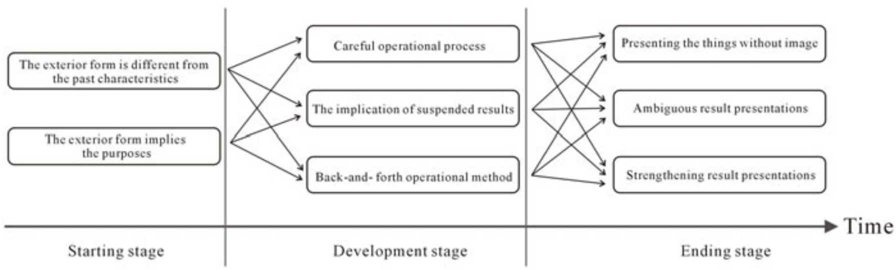


Fig. 5. Combinations of design techniques of different product-use stages

4 Conclusion and Discussion

This study is an exploratory investigation on user's expectation emotions and design techniques associated with product use. Using questionnaire, expectation case examples were collected. Through focus group discussion, the number of cases was reduced. A similarity degree evaluation and cluster analysis were used to further understand the expected image constructs. The starting stage of expectation is constructed by two sub-concepts: "attractive incentives" and "expectation on the results". The development stage of expectation is constructed by three sub-concepts: "cautious", "impatient" and "excited". The ending stage of expectation is constructed by three sub-concepts: "different from the expected", "going to be revealed" and "meeting the expectation".

Base on the expected image construct, a questionnaire was used to further investigate the product-use case examples for each sub-concept. After that, four design elements (exterior form, material presentation, function return, operational method) were used to evaluate the case examples. The evaluating result shows the key elements of starting stage are "exterior form" and "material presentation"; the key elements of development stage are "function return" and "operational method"; the key elements of ending stage are "function return" and "exterior form". A focus group discussion was held to generate the corresponding design techniques. According to the time sequence, there are 18 combinations of design techniques can be found.

This study targeted on the expectation emotion in 3 product-use stages, and generated the design techniques corresponding to the sub-concepts of expected image. However, a further study to evaluate and verify the effect of those generated design techniques can lead to a product associated with the expectation emotion is required.

References

1. Wilson, T.D., Klaaren, K.J.: Expectation whirls me round: The role of affective expectations on affective experiences. In: Clark, M.S. (ed.) Review of Personality and Social Psychology: Emotion and Social Behavior, vol. 14, pp. 1–31. Sage, Newbury Park (1992)
2. Plutchik, R.: A general psycho evolutionary theory of emotion, pp. 3–33 (1980)
3. Reed, E.S., Montgomery, M., Palmer, C., Pittenge, J.: Method for studying the invariant knowledge structure of action: conceptual organization of an everyday Action. The American Journal of Psychology 108, 37–65 (1995)
4. Tsao, Y.C., Hsiung, C.C.: Relationship between the adverbial images of complex actions in using utensils and the forms of the utensils. In: Proceedings of the 5th Annual International Conference on Industrial Engineering - Theory, Applications and Practice, HsinChu, Taiwan, December 13-15, pp. 577–582 (2000)
5. Chou, C.H.: Story, Wunan, Taipei (2002)
6. Tsao, Y.C., Chen, S.Y.: The expected image in the process of using products. In: Proceedings of the International Conference on Kansei Engineering and Emotion Research, Sapporo, Japan, October 10-12 (2007)

Designing the Personalized Nostalgic Emotion Value of a Product

Yu-Shan Tseng[1,2] and Ming-Chyuan Ho[1]

[1] Doctoral Program in Graduate School of Design
National Yunlin University of Science and Technology
[2] Kao Fong College of Digital Contents,
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
tsengyushan@gmail.com

Abstract. Personalization and nostalgic emotion are both human-centered. Human memory and emotions are the core concerns of design. It is important for a designer to involve consumers in the design process at very begging and communicate with them to integrate their emotional factors into design. This study discusses how to apply nostalgic emotional design strategies to personalize a product, grasp consumers' perception, and transform consumers' past memories into personalized exclusive symbols. This study adopts the qualitative approach, generalizes relevant literatures, makes verification and evaluation through practical design examples, and proposes the following suggestions: (1) using personal experience as the cut-in point in design; (2) role-playing of " story telling " (consumer side) and "listening to stories" (designer side); (3) integrating various modalities into design of a product. This study aims at constructing a tentative model for nostalgic-emotional design.

Keywords: personalized design, nostalgic.

1 Introduction

Consumers are gradually changing from rational consumption into emotional consumption and have an ever stronger subjective consciousness, and thus require diversified, differentiated and personalized products. Moreover, consumers' consumption mode is mostly affected by emotional feelings. Therefore, personalized design is a rewarding service model for innovative design. In recent years, nostalgic design technique has been found in various design fields like graphic design, advertisements, product design and drama. In terms of product design, it is a dynamic process influenced by many factors. The role of emotional issues in design procedures has been mentioned a lot recently. For instance, a series of design thinking methods concerning human and emotion have been highly valued and studied and applied into design and production processes.

1.1 Backgrounds and Motivations

Modern consumers not only pursue functionality and beauty, but also need products that can express their self-images. People create the imagery of "who they are" and

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 664–672, 2011.

their unique features through various manners such as clothes, foods, and cars or entertainment styles. These external substances favored by consumers reflect their self image, personality and taste. Consequently, personalized products appear in response to such consumption trend. Technique of nostalgic design, which also involves human emotion, endows product with vitality by means of story-telling or atmosphere creating. For elders with same experience, nostalgic design means being back to past memories; for young people, nostalgia is to understand the significance and value of old things. Personalized design offers consumers more chances of participation in design procedures. A professional designer should integrate various design factors and information resources, while excellent personalized design pursues mental resonance and emotional touch.

1.2 Research Purposes

Consumers are gradually changing from rational consumption into emotional consumption and have an ever stronger subjective consciousness, and thus require diversified, differentiated and personalized products. Personalized design exactly adopts the product design strategy of differentiation. As we pursue products that meet spiritual demands rather than mass products, present-day merchandise should not only have functions but also can help consumers present their special taste, preference, personality and difference from others. Design that satisfies individual consumption requirement can show consumer's difference, self characteristic and style so as to realize his/her idea of "pursuing differences while seeking common ground". To discuss the application of nostalgic design into personalized product design, this study investigates consumers' perception and converts their past memories into personalized exclusive symbols. Qualitative research, literature review and case discussion are used in the current paper. Purposes of this study are as follows:

1. Relevant literatures are collected to study the feasibility of combining nostalgic emotion with personalized product.
2. This study discusses about the advantages and shortcomings of design techniques of nostalgic products available on the current market.
3. It generalizes the methods of nostalgic design so as to provide references for designers who implement personalized product design.

2 Relevant Researches

Human have emotions. For things in the past, people, matter or thing, they have memories, complicated feelings and emotions in various degrees. Nostalgic design aims to awaken our deep emotional reactions. Nostalgia is a kind of nostalgic phenomenon valued and discussed by various research fields such as history, psychology, sociology, humanics, environmental psychology and other social sciences. It is also extensively used in the area of design.

2.1 Nostalgic Emotion

Belk (1990) suggests that nostalgia is to devote to emotion through memory and retrospection and represents a person's hobby of collection, adoration of old things or preservation of wonderful past feeling. Sociologist Davis (1979) believes the significance of nostalgia lies in having a vision of the past. Stern (1992) defines nostalgia as a state of individual mood. The content of memory is idealized and extracts nice memory from past times. Curio and object collection are good examples. Ackbar Abbas (1997) thinks nostalgia is pulling people's memory back to the past. People of different generations are quite different in nostalgic characteristics and value proposition: people born in 1940s~1950s have nostalgia for events affected by external environment, so their nostalgia is mainly public-oriented style and individual-oriented nostalgia is rare; those born in 1960s mostly have public nostalgia oriented by individual experience and public events; people of 1970s have individual nostalgia emphasized on personal experience and events and seldom involve in public nostalgia; nowadays, young people born in 1980s are inclined to exploratory and entertainment-style public nostalgia (Kuo, C.L., et al., 2009).

Therefore, we know that nostalgia is about life experience and memory and it awakens our emotional memories toward past people, events and things in the past. Nostalgia involves both pleasant and unpleasant moods. Nostalgia, which refers to a vision of the past, is idealized and unpractical and usually relevant to a warm childhood memory, a certain game or precious personal items. Classified discussions about nostalgia proposed by relevant scholars are summarized in the following table:

Table 1 Nostalgia Types. (summarized by this study)

Scholars	Nostalgia Types
Davis(1979)	(1) Simple Nostalgia (2) Reflexive Nostalgia (3) Interpretive Nostalgia
Stern(1992)	(1) Personal Nostalgia (2) Historical Nostalgia
Baker & Kennedy (1994)	(1) Real Nostalgia (2) Simulated Nostalgia (3) Collective Nostalgia
Holak & Havlena (1997)	(1) Personal (2) Culture (3) Inter personal (4) Virtual

These scholars classify nostalgia into two kinds: one is mainly affected by individual experience, which belongs to direct experience; the other is nostalgia for history, which belongs to the public and group.

2.2 Personalized Design

Personalized products can meet consumers' mental demands, e.g. "esteem needs" and "self-actualization needs" contained in Maslow's Needs-Hierarchy Theory. The significance of personalized product lies in highlighting differences between products and symbolizing the unique mark of product..As mass- customization design evolves into personalized design, manufacturing line model of old industry also gradually changes into product design mode that emphasizes individual style. In the past, we pursued goods with excellent quality and reasonable price; nowadays, we need "personalized" products (Lin, R.T, 2005).

When designing personalized products, through making interactive decision about the shape, material, color, voice, texture, etc of product, designers can create a kind of specified product that makes consumers identify its feature(Govers, Hekkert, and Schoormans 2002; Jordan 2002). Solomon (1983) also points that a product appearance helps shape the image of the users Traditional design method strives for commonness of design, and designers utilize design experience to achieve product uniformity and popularization. However, the generation of personalized product is consumer-centered, analyzes consumer requirement, and design personalized product according to consumers' individualized needs. Personalized product has a feature of somewhat showing the owner's characteristics like social status, economic status, and personal preference and interest.

2.3 Personalized Design and Nostalgic Emotion

Currently, it is a trend to integrate nostalgic style into design and recover the classic style. While personalized design is produced because people need a more obvious personal symbol to show their distinctness. As Norman (2005) indicated, when we are exposed to a single thing, our reaction is not only influenced by its usability but also dependent on if it actually reflects our image or awakens our nostalgia. Such emotion can be evoked by vision, hearing, smell, touch and taste. These familiar voices and flavors stimulate our nostalgic emotions. Nostalgic design is common in marketing, advertisement, product design, etc, and the most frequently used technique is "story telling", which converts nostalgic materials into story and causes consumers' collective nostalgia and resonance. Nostalgic emotion is originated from personal memory or some people's joint memory. Nostalgic design aims to awaken consumers' memory of past times, re-interpret product appearance and pattern from the angle of design, as well as lead consumers to go back to the past by time machine and re-experience nice old times. The nostalgic emotional value of personalized goods can achieve consumers' emotional satisfaction and mental identification.

In terms of three design hierarchies proposed by Norman(2005), this study classifies nostalgic design and personalized design into reflective level because design of this level emphasizes on expressing individual image and memories of old experiences. Moreover, nostalgic moods can endow objects with symbolic meaning; while mass production model values the design of appearance and function (seen in table 2). When integrating nostalgic technique into personalized design, designers should not only know consumers' style, needs and preferences, but also abstract elements from consumers' direct or indirect nostalgic memories during design procedures, analyze the collected data as well as use familiar things for determining main design elements and formulating design strategy.

Table 2. Diagram about relationships between three product design levels and nostalgic personalized design. (Summarized by this study).

Three levels of design	Product characteristics	Method of product manufacturing
Visceral	Appearance design which is about how things look, feel, and sound.	Mass production
Behavioral	Behavioral design is emphasis on function which is about product function well and easily accessible	Mass production
Reflective	Reflective design is about the meaning of things, user's emotion or memory, Self- image and use pleasure.	Nostalgic design Personalized design

2.4 Design Case Analysis and Discussion

Through analyzing and generalizing the products of current nostalgic design, this study investigates the application of nostalgic design strategy into products, and the relevance between product and personalized design. Relevant design examples are summarized as shown in Table 3:

Table 3. Cases about the application of nostalgic emotion into product design

Product picture	Nostalgic design strategy	Personalized design
 Sources: http://www.makeamixa.com	Nostalgic cassette shape is combined with the function of USB and both posses an imagery of storing. The product not only has functionality but also makes consumers think of old-time fineness.	It belongs to customized design. Consumers can upload their own pictures or utilize simple mapping interface to implement participatory design in accordance with colors and figures provided by the website.

Table 3. (*Continued*)

 Sources: http://www.eslite.com	It is a kind of collective-nostalgic memory. Figurative technique is adopted to express rocking horse, with which most people played in childhood.	Only two kinds of leather in different colors are available, and there's no personalized design.
 Sources: http://www.magno-design.com	Nostalgic radio made of ebony presents people's collective nostalgic emotions	Handmade product. The appearance shape and operation mode are designed and manufactured following classic style. There's no personalized design aimed at consumer requirement.
 Sources: http://www.bluemic.com/	Classic-style USB microphone evokes people's nostalgic emotion.	Figurative design technique. There's no personalized design aimed at consumer requirement.
 Sources: http://www.electrojoe.co.uk/	Audio cassette-shaped USB hub. Nostalgic product and different functions.	Figurative design technique. There's no personalized design aimed at consumer requirement.

Table 3. (*Continued*)

 Sources: http://www.nvdrsdesign.com/	People who have experienced the times of music cassette must be familiar with this. Once the product is out of power, you can generate electricity through turning around the holes in the cassette with your finger. This reminds people of adjusting magnetic tape by fingers in the past.	Although directly applying nostalgic shape, in terms of the design of user interface, it realizes nostalgic effects through operating method familiar to consumers. There's no personalized design aimed at consumer requirement.
 Sources: http://www.burakkaynak.com/	This product uses 3.5 inch floppy disk to change into a function of sticky notes, and has nostalgic value.	It transfers the product functionality through figurative nostalgic shape. There's no personalized design aimed at consumer requirement.
 Sources: http://www.privatecircle.it	In the era without CD and MP3 products, "cassette" is the main media for pop music. The cassette wallet awakens people's memory of that time.	This kind of design redesigns the discarded cassette and endows the product a new image and function so that nostalgia can be continued in another form.

3 Application of Nostalgic Emotion into Personalized Design

During the processes of personalized design, only through knowing consumer requirement can designers endow product with unique feature, communication capacity and emotional interaction with users. In other words, modern product design should be a kind of design activity involving people's emotion rather than just a process of shape modeling as the initial stage of design development. If a designer can understand consumers' nostalgia for human, event, object, etc., and integrate representative nostalgic features into design, they will produce personalized design that values the strategy of differentiation; this kind of design can better meet consumers' demand on expressing personal style and taste and realize their goal of being different from others.

Currently, the expressive technique of nostalgic products is mostly the replication and application of old shape, probably because a direct imitation of appearance shape can attract consumers' attention more easily. This kind of nostalgic design lacks individual nostalgic emotion and experience, so it is collective nostalgia and inapplicable to personalized design. For integrating nostalgic emotions into personalized design, this paper proposes the following suggestions:

1. Person's direct past experience as the basis of design- nostalgia is sorted into individual nostalgia and collective nostalgia. Individual nostalgia is the most direct experience, and past memory and experience is the basis of people's differences from others. Features of individual nostalgia are applicable to personalized design.
2. Role play of telling story and listening to story- when applying nostalgic design strategy into personalized design, the role of the consumer is to tell story and the designer is the person who listens to the story. Then, the designer should decompose the story into design elements and convert the story into product possessing symbolism and distinctness.
3. The application of nostalgic emotion into five-sense metaphorical design- five senses refer to vision, touch, hearing, smell and taste. The external shape affected by vision is a kind of direct application into design, which utilizes color, figure, pattern, etc. Moreover, voice, odor, touch or flavor familiar to individuals can also be integrated into expressive elements of personalized product design.

4 Conclusions and Suggestions

This is an era of emotional consumption. When purchasing goods, consumers don't emphasize on quantity, quality and price any more, but pursue emotional satisfaction and mental identification. Personalization and nostalgic emotion are both human centered, and consider people's memory and emotion as the core of design. Personalized nostalgic products start from consumers' emotional needs, evoke their precious memories and induce their spiritual resonance, so as to meet consumer requirement. Consumers' participation in design can integrate their individual emotion and preference, and realize their demands with designers' assistance in design. The role of designers is to obtain usable information through their interactions and communications with consumers, integrate elements (e.g. shape, symbol, color) that symbolize consumers' personality and design products that feature distinctness and represent consumers' characteristics. If nostalgic elements relevant to personal emotion and memory are integrated into the design strategy, personalized products designed by designers will be more effective and touch people's inner hearts. 21st century has become an era when consumers pursue personalized design and enterprises actively adopt the strategy of emotional marketing. A more effective design strategy is to meet consumers' demand on personalized products and develop goods that can stimulate emotional resonance. For future researches, the current paper suggests having in-depth interviews with designers or experts, finding out their method for choosing elements of personalized design, and making consequent verification through comparing design results with consumers' expected effects.

References

1. Abbas, A.: Hong Kong: culture and the politics of disappearance, Minneapo lis University of Minnesota Press (1997)
2. Belk, R.W.: The role of possessions in constructing and maintaining a sense of past. Advances in Consumer Research 17, 669–676 (1990)
3. Baker, S.M., Kennedy, P.F.: Death By Nostalgia: A Diagnosis of context-specific Cases. Advances in Consumer Research 21, 169–174 (1994)
4. Davis, F.: Yearning for yesterday: sociology of nostalgia. Free press, New York (1979)
5. Norman, D.A.: Emotional Design: Why We Love (or Hate) Everyday Things. Baker & Taylor, New York (2005)
6. Huppatz, D.J.: Designer Nostalgia in Hong Kong. Design Issues 25, 14–28 (2009)
7. Govers, P.C.M., Hekkert, P., Schoormans, J.P.L.: Happy, Cute and Tough: Can Designers Create a Product Personality That Consumers Understand? In: McDonagh, D., Hekkert, P., Van Erp, J., Gyi, D. (eds.) Design and Emotion: The Experience of Everyday Things, pp. 345–349. Taylor and Francis, London (2002)
8. Holak, S.L., Havlena, W.J.: Feelings, fantasies, and memories: An examination of the emotional components of nostalgia. Journal of Business Research 42, 217–226 (1998)
9. Jordan, P.W.: The Personalities of Products. In: Green, W.S., Jordan, P.W. (eds.) Pleasure with Products: Beyond Usability, pp. 19–47. Taylor & Francis, London (2002)
10. Kuo, C.L., Yeh, H.C., Li, C.C., Tang, H.C., Wu, M.Y.: Create a happy new life: Nostalgia service opportunities and practice. Industrial Economics & Knowledge Center, IEK, Taipei (2009)
11. Solomon, M.R.: The Role of Products as Social Stimuli: A Symbolic Interactionism Pers pective. Journal of Consumer Research 10, 319–329 (1983)
12. Stern, B.B.: Nostalgia in advertising text: Romancing the past. Advances in Consumer Research 19, 388–389 (1999)
13. Lin, R.T.: Combination of Technology and Humanity-Cultural Creativity. Science Development 396, 68–75 (2005)
14. Magnetic North Interactive Ltd, `http://www.makeamixa.com`
15. Eslite.com, `http://www.eslite.com`
16. Kartono, S.S.: `http://www.magno-design.com`
17. Blue Microphones, `http://www.bluemic.com/`
18. Foschi, M.: `http://www.privatecircle.it/`
19. Kaynak, B.: `http://www.burakkaynak.com/`
20. Joe, E.: `http://www.electrojoe.co.uk/`
21. NVDRS design, `http://www.nvdrsdesign.com/`

Author Index