Lecture Notes in Computer Science 4559

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, UK

Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler
University of Surrey, Guildford, UK

Jon M. Kleinberg
Cornell University, Ithaca, NY, USA

Friedemann Mattern
ETH Zurich, Switzerland

John C. Mitchell
Stanford University, CA, USA

Moni Naor
Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz
University of Bern, Switzerland

C. Pandu Rangan
Indian Institute of Technology, Madras, India

Bernhard Steffen
University of Dortmund, Germany

Madhu Sudan
Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Moshe Y. Vardi
Rice University, Houston, TX, USA

Gerhard Weikum
Max-Planck Institute of Computer Science, Saarbruecken, Germany

Nuray Aykin (Ed.)

Usability and Internationalization

HCI and Culture

Second International Conference on
Usability and Internationalization, UI-HCII 2007
Held as Part of HCI International 2007
Beijing, China, July 22-27, 2007
Proceedings, Part I

Springer

Volume Editor

Nuray Aykin
The New School
55 West 13th Street, New York, NY 10011, USA
E-mail: aykinn@newschool.edu

Library of Congress Control Number: 2007929579

CR Subject Classification (1998): H.5.2, H.5.3, H.3-5, C.2, K.4, D.2, K.6

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

ISSN 0302-9743
ISBN-10 3-540-73286-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-73286-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12082605 06/3180 5 4 3 2 1 0

Foreword

The 12th International Conference on Human-Computer Interaction, HCI International 2007, was held in Beijing, P.R. China, 22-27 July 2007, jointly with the Symposium on Human Interface (Japan) 2007, the 7th International Conference on Engineering Psychology and Cognitive Ergonomics, the 4th International Conference on Universal Access in Human-Computer Interaction, the 2nd International Conference on Virtual Reality, the 2nd International Conference on Usability and Internationalization, the 2nd International Conference on Online Communities and Social Computing, the 3rd International Conference on Augmented Cognition, and the 1st International Conference on Digital Human Modeling.

A total of 3403 individuals from academia, research institutes, industry and governmental agencies from 76 countries submitted contributions, and 1681 papers, 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 Nuray Aykin, contains papers in the thematic area of Usability and Internationalization, addressing the following major topics:

- Cross-Cultural Design
- International and Intercultural Usability
- User Studies

The remaining volumes of the HCI International 2007 proceedings are:

- Volume 1, LNCS 4550, Interaction Design and Usability, edited by Julie A. Jacko
- Volume 2, LNCS 4551, Interaction Platforms and Techniques, edited by Julie A. Jacko
- Volume 3, LNCS 4552, HCI Intelligent Multimodal Interaction Environments, edited by Julie A. Jacko
- Volume 4, LNCS 4553, HCI Applications and Services, edited by Julie A. Jacko
- Volume 5, LNCS 4554, Coping with Diversity in Universal Access, edited by Constantine Stephanidis
- Volume 6, LNCS 4555, Universal Access to Ambient Interaction, edited by Constantine Stephanidis
- Volume 7, LNCS 4556, Universal Access to Applications and Services, edited by Constantine Stephanidis
- Volume 8, LNCS 4557, Methods, Techniques and Tools in Information Design, edited by Michael J. Smith and Gavriel Salvendy
- Volume 9, LNCS 4558, Interacting in Information Environments, edited by Michael J. Smith and Gavriel Salvendy
- Volume 11, LNCS 4560, Global and Local User Interfaces, edited by Nuray Aykin

- Volume 12, LNCS 4561, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 13, LNAI 4562, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris
- Volume 14, LNCS 4563, Virtual Reality, edited by Randall Shumaker
- Volume 15, LNCS 4564, Online Communities and Social Computing, edited by Douglas Schuler
- Volume 16, LNAI 4565, Foundations of Augmented Cognition 3rd Edition, edited by Dylan D. Schmorrow and Leah M. Reeves
- Volume 17, LNCS 4566, Ergonomics and Health Aspects of Work with Computers, edited by Marvin J. Dainoff

I would like to thank the Program Chairs and the members of the Program Boards of all Thematic Areas, listed below, for their contribution to the highest scientific quality and the overall success of the HCI International 2007 Conference.

Ergonomics and Health Aspects of Work with Computers

Program Chair: Marvin J. Dainoff

Arne Aaras, Norway
Pascale Carayon, USA
Barbara G.F. Cohen, USA
Wolfgang Friesdorf, Germany
Martin Helander, Singapore
Ben-Tzion Karsh, USA
Waldemar Karwowski, USA
Peter Kern, Germany
Danuta Koradecka, Poland
Kari Lindstrom, Finland
Holger Luczak, Germany
Aura C. Matias, Philippines
Kyung (Ken) Park, Korea
Michelle Robertson, USA
Steven L. Sauter, USA
Dominique L. Scapin, France
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

Lajos Balint, Hungary
Gunilla Bradley, Sweden
Hans-Jörg Bullinger, Germany
Alan H.S. Chan, Hong Kong
Klaus-Peter Fähnrich, Germany
Michitaka Hirose, Japan
Yoshinori Horie, Japan
Richard Koubek, USA
Yasufumi Kume, Japan
Mark Lehto, USA
Jiye Mao, P.R. China
Robert Proctor, USA
Youngho Rhee, Korea
Anxo Cereijo Roibás, UK
Francois Sainfort, USA
Katsunori Shimohara, Japan
Tsutomu Tabe, Japan
Alvaro Taveira, USA
Kim-Phuong L. Vu, USA
Tomio Watanabe, Japan
Sakae Yamamoto, Japan
Hidekazu Yoshikawa, Japan

Fiona Nah, USA
Shogo Nishida, Japan
Leszek Pacholski, Poland
Li Zheng, P.R. China
Bernhard Zimolong, Germany

Human-Computer Interaction

Program Chair: Julie A. Jacko

Sebastiano Bagnara, Italy
Jianming Dong, USA
John Eklund, Australia
Xiaowen Fang, USA
Sheue-Ling Hwang, Taiwan
Yong Gu Ji, Korea
Steven J. Landry, USA
Jonathan Lazar, USA
V. Kathlene Leonard, USA
Chang S. Nam, USA
Anthony F. Norcio, USA
Celestine A. Ntuen, USA
P.L. Patrick Rau, P.R. China
Andrew Sears, USA
Holly Vitense, USA
Wenli Zhu, P.R. China

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

Kenneth R. Boff, USA
Guy Boy, France
Pietro Carlo Cacciabue, Italy
Judy Edworthy, UK
Erik Hollnagel, Sweden
Kenji Itoh, Japan
Peter G.A.M. Jorna, The Netherlands
Kenneth R. Laughery, USA
Nicolas Marmaras, Greece
David Morrison, Australia
Sundaram Narayanan, USA
Eduardo Salas, USA
Dirk Schaefer, France
Axel Schulte, Germany
Neville A. Stanton, UK
Andrew Thatcher, South Africa

Universal Access in Human-Computer Interaction

Program Chair: Constantine Stephanidis

Julio Abascal, Spain
Ray Adams, UK
Elizabeth Andre, Germany
Margherita Antona, Greece
Chieko Asakawa, Japan
Christian Bühler, Germany
Noelle Carbonell, France
Jerzy Charytonowicz, Poland
Pier Luigi Emiliani, Italy
Michael Fairhurst, UK
Gerhard Fischer, USA
Zhengjie Liu, P.R. China
Klaus Miesenberger, Austria
John Mylopoulos, Canada
Michael Pieper, Germany
Angel Puerta, USA
Anthony Savidis, Greece
Andrew Sears, USA
Ben Shneiderman, USA
Christian Stary, Austria
Hirotada Ueda, Japan
Jean Vanderdonckt, Belgium

Jon Gunderson, USA
Andreas Holzinger, Austria
Arthur Karshmer, USA
Simeon Keates, USA
George Kouroupetroglou, Greece
Jonathan Lazar, USA
Seongil Lee, Korea
Gregg Vanderheiden, USA
Gerhard Weber, Germany
Harald Weber, Germany
Toshiki Yamaoka, Japan
Mary Zajicek, UK
Panayiotis Zaphiris, UK

Virtual Reality

Program Chair: Randall Shumaker

Terry Allard, USA
Pat Banerjee, USA
Robert S. Kennedy, USA
Heidi Kroemker, Germany
Ben Lawson, USA
Ming Lin, USA
Bowen Loftin, USA
Holger Luczak, Germany
Annie Luciani, France
Gordon Mair, UK
Ulrich Neumann, USA
Albert "Skip" Rizzo, USA
Lawrence Rosenblum, USA
Dylan Schmorrow, USA
Kay Stanney, USA
Susumu Tachi, Japan
John Wilson, UK
Wei Zhang, P.R. China
Michael Zyda, USA

Usability and Internationalization

Program Chair: Nuray Aykin

Genevieve Bell, USA
Alan Chan, Hong Kong
Apala Lahiri Chavan, India
Jori Clarke, USA
Pierre-Henri Dejean, France
Susan Dray, USA
Paul Fu, USA
Emilie Gould, Canada
Sung H. Han, South Korea
Veikko Ikonen, Finland
Richard Ishida, UK
Esin Kiris, USA
Tobias Komischke, Germany
Masaaki Kurosu, Japan
James R. Lewis, USA
Rungtai Lin, Taiwan
Aaron Marcus, USA
Allen E. Milewski, USA
Patrick O'Sullivan, Ireland
Girish V. Prabhu, India
Kerstin Röse, Germany
Eunice Ratna Sari, Indonesia
Supriya Singh, Australia
Serengul Smith, UK
Denise Spacinsky, USA
Christian Sturm, Mexico
Adi B. Tedjasaputra, Singapore
Myung Hwan Yun, South Korea
Chen Zhao, P.R. China

Online Communities and Social Computing

Program Chair: Douglas Schuler

Chadia Abras, USA
Lecia Barker, USA
Amy Bruckman, USA
Peter van den Besselaar, The Netherlands
Peter Day, UK
Fiorella De Cindio, Italy
John Fung, P.R. China
Michael Gurstein, USA
Tom Horan, USA
Piet Kommers, The Netherlands
Jonathan Lazar, USA
Stefanie Lindstaedt, Austria
Diane Maloney-Krichmar, USA
Isaac Mao, P.R. China
Hideyuki Nakanishi, Japan
A. Ant Ozok, USA
Jennifer Preece, USA
Partha Pratim Sarker, Bangladesh
Gilson Schwartz, Brazil
Sergei Stafeev, Russia
F.F. Tusubira, Uganda
Cheng-Yen Wang, Taiwan

Augmented Cognition

Program Chair: Dylan D. Schmorrow

Kenneth Boff, USA
Joseph Cohn, USA
Blair Dickson, UK
Henry Girolamo, USA
Gerald Edelman, USA
Eric Horvitz, USA
Wilhelm Kincses, Germany
Amy Kruse, USA
Lee Kollmorgen, USA
Dennis McBride, USA
Jeffrey Morrison, USA
Denise Nicholson, USA
Dennis Proffitt, USA
Harry Shum, P.R. China
Kay Stanney, USA
Roy Stripling, USA
Michael Swetnam, USA
Robert Taylor, UK
John Wagner, USA

Digital Human Modeling

Program Chair: Vincent G. Duffy

Norm Badler, USA
Heiner Bubb, Germany
Don Chaffin, USA
Kathryn Cormican, Ireland
Andris Freivalds, USA
Ravindra Goonetilleke, Hong Kong
Anand Gramopadhye, USA
Sung H. Han, South Korea
Pheng Ann Heng, Hong Kong
Dewen Jin, P.R. China
Kang Li, USA
Zhizhong Li, P.R. China
Lizhuang Ma, P.R. China
Timo Maatta, Finland
J. Mark Porter, UK
Jim Potvin, Canada
Jean-Pierre Verriest, France
Zhaoqi Wang, P.R. China
Xiugan Yuan, P.R. China
Shao-Xiang Zhang, P.R. China
Xudong Zhang, USA

In addition to the members of the Program Boards above, I also wish to thank the following volunteer external reviewers: Kelly Hale, David Kobus, Amy Kruse, Cali Fidopiastis and Karl Van Orden from the USA, Mark Neerincx and Marc Grootjen from the Netherlands, Wilhelm Kincses from Germany, Ganesh Bhutkar and Mathura Prasad from India, Frederick Li from the UK, and Dimitris Grammenos, Angeliki Kastrinaki, Iosif Klironomos, Alexandros Mourouzis, and Stavroula Ntoa from Greece.

This conference could not have been possible without the continuous support and advise of the Conference Scientific Advisor, Prof. Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications Chair and Editor of HCI International News, Abbas Moallem, and of the members of the Organizational Board from P.R. China, Patrick Rau (Chair), Bo Chen, Xiaolan Fu, Zhibin Jiang, Congdong Li, Zhenjie Liu, Mowei Shen, Yuanchun Shi, Hui Su, Linyang Sun, Ming Po Tham, Ben Tsiang, Jian Wang, Guangyou Xu, Winnie Wanli Yang, Shuping Yi, Kan Zhang, and Wei Zho.

I would also like to thank for their contribution towards the organization of the HCI International 2007 Conference the members of the Human Computer Interaction Laboratory of ICS-FORTH, and in particular Margherita Antona, Maria Pitsoulaki, George Paparoulis, Maria Bouhli, Stavroula Ntoa and George Margetis.

Constantine Stephanidis
General Chair, HCI International 2007

HCI International 2009

The 13th International Conference on Human-Computer Interaction, HCI International 2009, will be held jointly with the affiliated Conferences in San Diego, California, USA, in the Town and Country Resort & Convention Center, 19-24 July 2009. It will cover a broad spectrum of themes related to Human Computer Interaction, 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. For more information, please visit the Conference website: http://www.hcii2009.org/

General Chair
Professor Constantine Stephanidis
ICS-FORTH and University of Crete
Heraklion, Crete, Greece
Email: program@hcii2009.org

Table of Contents

Part I: Cross-Cultural Design

Panel Discussion: Global Innovative Design for Social Change 3
Nuray Aykin, Apala Lahiri Chavan, Susan M. Dray, and Girish Prabhu

Enabling User Centered Design Processes in Open Source Communities 10
Mads Bødker, Lene Nielsen, and Rikke N. Orngreen

A Dramatic Day in the Life of a Shared Indian Mobile Phone 19
Apala Lahiri Chavan

Smart Strategies for Creating Culture Friendly Products and Interfaces 27
Apala Lahiri Chavan

When in Rome... Be Yourself: A Perspective on Dealing with Cultural Dissimilarities in Ethnography 33
Apala Lahiri Chavan and Rahul Ajmera

Designing User Interfaces for Mobile Entertaining Devices with Cross-Cultural Considerations 37
Chien-Hsiung Chen and Chia-Ying Tsai

Kansei Design with Cross Cultural Perspectives 47
Kuohsiang Chen, Shu-chuan Chiu, and Fang-chyuan Lin

The Challenge of Dealing with Cultural Differences in Industrial Design in Emerging Countries: Latin-American Case Studies 57
Alvaro Enrique Diaz

Emerging Issues in Doing Cross-Cultural Research in Multicultural and Multilingual Societies 65
Henry Been-Lirn Duh and Vivian Hsueh-Hua Chen

The Digital and the Divine: Taking a Ritual View of Communication and ICT Interaction 74
Brooke Foucault and Jay Melican

Shanghaied in a User-Friendly Manner - An American's Initial Experiences in a Full-Time Usability Job in China 83
Brian I. Glucroft

A Tool for Cross-Cultural Human Computer Interaction Analysis 89
Rüdiger Heimgärtner

Locating Culture in HCI with Information Kiosks and Social Networks 99
Tom Hope, Masahiro Hamasaki, Keisuke Ishida, Noriyuki Fujimura, Yoshiyuki Nakamura, and Takuichi Nishimura

HCI and SE – The Cultures of the Professions 108
Joshi Anirudha

Development of Integrated Analysis System and Tool of Perception, Recognition, and Behavior for Web Usability Test: With Emphasis on Eye-Tracking, Mouse-Tracking, and Retrospective Think Aloud 113
Byungjoo Kim, Ying Dong, Sungjin Kim, and Kun-Pyo Lee

Cultural Difference and Its Effects on User Research Methodologies 122
Jungjoo Lee, Thu-Trang Tran, and Kun-Pyo Lee

A Development of Graphical Interface for Decision Making Process Including Real-Time Consistency Evaluation 130
Joong-Ho Lee, Ki-Won Yeom, and Ji-Hyung Park

Using Webzine to Create Effective Communications Between China and the West 138
Christina Li, Sean Liu, and Eleanor Lisney

Designing "Culture" into Modern Product: A Case Study of Cultural Product Design 146
Rungtai Lin, Ming-Xian Sun, Ya-Ping Chang, Yu-Ching Chan, Yi-Chen Hsieh, and Yuan-Ching Huang

Digital Archive Database for Cultural Product Design 154
Rungtai Lin, Ricer Cheng, and Ming-Xian Sun

Cross-Cultural Understanding of Content and Interface in the Context of E-Learning Systems 164
Abdalghani Mushtaha and Olga De Troyer

Differences in Task Descriptions in the Think Aloud Test 174
Lene Nielsen and Sameer Chavan

The Use of Cognitive and Social Psychological Principles in Field Research: How It Furthers Our Understanding of User Behaviors, Needs and Motivations, and Informs the Product Design Process 181
Krisela Rivera and Elissa Darnell

The Role of Annotation in Intercultural Communication 186
Tomohiro Shigenobu, Kunikazu Fujii, and Takashi Yoshino

An Activity Approach to Cross-Cultural Design 196
Huatong Sun

Creating an International Design Team 206
Becky Sundling

Incorporating the Cultural Dimensions into the Theoretical Framework of Website Information Architecture 212
Wan Abdul Rahim Wan Mohd Isa, Nor Laila Md Noor, and Shafie Mehad

Part II: International and Intercultural Usability

Cross-Use: Cross-Cultural Usability User Evaluation- In-Context 225
Jasem M. Alostath and Abdulwahed Moh Khalfan

Testing Remote Users: An Innovative Technology 235
Rebecca Matson Sukach Baker, Esin Kiris, and Omar Vasnaik

Web Usability and Evaluation: Issues and Concerns 243
S. Batra and R.R. Bishu

The Impact of Different Icon Sets on the Usability of a Word Processor 250
Tanya R. Beelders, P.J. Blignaut, T. McDonald, and E. Dednam

Systems Development Methods and Usability in Norway: An Industrial Perspective 258
Bendik Bygstad, Gheorghita Ghinea, and Eivind Brevik

Activities for Usability in Lenovo China 267
Baihong Chen and Rong Yang

The Cultural Usability (CULTUSAB) Project: Studies of Cultural Models in Psychological Usability Evaluation Methods 274
Torkil Clemmensen and Tom Plocher

Cultural Usability Tests – How Usability Tests Are Not the Same All over the World 281
Torkil Clemmensen, Qingxin Shi, Jyoti Kumar, Huiyang Li, Xianghong Sun, and Pradeep Yammiyavar

Getting the Most Out of Personas for Product Usability Enhancements 291
Jianming Dong, Kuldeep Kelkar, and Kelly Braun

Testing Object Management (TOM): A Prototype for Usability Knowledge Management in Global Software 297
Ian Douglas

Assessing Usability Problems in Latin-American Academic Webpages with Cognitive Walkthroughs and Datamining Techniques 306
María Paula González, Jesús Loréss, and Antoni Granollers

Usability Constructs: A Cross-Cultural Study of How Users and Developers Experience Their Use of Information Systems 317
Morten Hertzum, Torkil Clemmensen, Kasper Hornbæk, Jyoti Kumar, Qingxin Shi, and Pradeep Yammiyavar

A Study for Usability Risk Level in Physical User Interface of Mobile Phone 327
Beomsuk Jin, Sangmin Ko, Jaeseung Mun, and Yong Gu Ji

Tracing Cognitive Processes for Usability Evaluation: A Cross Cultural Mind Tape Study 336
Jyoti Kumar, Janni Nielsen, and Pradeep Yammiyavar

Lessons from Applying Usability Engineering to Fast-Paced Product Development Organizations 346
Dong-Seok Lee and Young-Hwan Pan

An Axiomatic Method for Cross Cultural Usability Analysis 355
Sheau-Farn Max Liang

The Impact of Culture on Usability: Designing Usable Products for the International User 365
Carol Lodge

A Digital Training System for Freehand Sketch Practice 369
Ding-Bang Luh and Shao-Nung Chen

Culture Issues in Traffic Sign Usability 379
Annie W.Y. Ng and Alan H.S. Chan

International Remote Usability Evaluation: The Bliss of Not Being There 388
Mika P. Nieminen, Petri Mannonen, and Johanna Viitanen

A Framework for Evaluating the Usability of Spoken Language Dialog Systems (SLDSs) 398
Wonkyu Park, Sung H. Han, Yong S. Park, Jungchul Park, and Huichul Yang

Usability of Adaptable and Adaptive Menus 405
Jungchul Park, Sung H. Han, Yong S. Park, and Youngseok Cho

Towards Detecting Cognitive Load and Emotions in Usability Studies Using the RealEYES Framework 412
Randolf Schultz, Christian Peter, Michael Blech, Jörg Voskamp, and Bodo Urban

Relationship Model in Cultural Usability Testing 422
Qingxin Shi and Torkil Clemmensen

An Empirical Evaluation of Graphical Usable Interface on Mobile Chat 432
Victoria Yee Siew Yen and Daniel Su Kuen Seong

A Tale of Two Teams: Success and Failure in Virtual Team Meetings . . . 442
Marilyn M. Tremaine, Allen Milewski, Richard Egan, and Suling Zhang

Assumptions Considered Harmful 452
Heike Winschiers and Jens Fendler

Analyzing Non-verbal Cues in Usability Evaluation Tests 462
Pradeep Yammiyavar, Torkil Clemmensen, and Jyoti Kumar

Online Analysis of Hierarchical Events in Meetings 472
Xiang Zhang, Guang-You Xu, Xiao-Ling Xiao, and Lin-Mi Tao

Part III: User Studies

A Cross Culture Study on Phone Carrying and Physical Personalization 483
Yanqing Cui, Jan Chipchase, and Fumiko Ichikawa

Performance Modeling Using Anthropometry for Minority Population 493
V. Gnaneswaran and R.R. Bishu

Investigating the Differences in Web Browsing Behaviour of Chinese and European Users Using Mouse Tracking 502
Lee Griffiths and Zhongming Chen

The Effect of Morphological Elements on the Icon Recognition in Smart Phones 513
Chiwu Huang and Chieh-Ming Tsai

Performance Evaluation of the Wheel Navigation Key Used for Mobile Phone and MP3 523
Hyun-Wook Jung and Jung-Yong Kim

Correlation Between Cognitive Style and Structure and Flow in Mobile Phone Interface: Comparing Performance and Preference of Korean and Dutch Users 531
Ji Hye Kim, Kun-Pyo Lee, and Im Kyeong You

Incorporating JND into the Design of Mobile Device Display 541
Joo Hwan Lee, Won Yong Suh, Cheol Lee, Jang Hyeon Jo, and Myung Hwan Yun

Fit Evaluation of 3D Virtual Garment 550
Joohyun Lee, Yunja Nam, Ming Hai Cui, Kueng Mi Choi, and Young Lim Choi

Evaluation of Two Pointing Control Devices for a Cellular Phone 559
Ji Hyoun Lim, Cheol Lee, Sun Young Park, and Myung Hwan Yun

Design and Evaluation of a Handled Trackball as a Robust Interface in Motion 566
Chiuhsiang Joe Lin, Chi-No Liu, and Jun-Lung Hwang

Impact of Culture on International User Research -A Case Study: Integration Pre-study in Paper Mills 576
Anna Oikarinen and Marko Nieminen

Computer Mediated Banking: A Cross-Cultural Analysis of SMEs 586
Alison Ruth and Jenine Beekhuyzen

A Comparative Study of Thai and UK Older Web Users 596
Prush Sa-nga-ngam and Sri Kurniawan

A Qualitative Oriented Study About IT Procurement Processes: Comparison of 4 European Countries 606
Michael Schiessl and Sabrina Duda

An Empirical Study on the Smallest Comfortable Button/Icon Size on Touch Screen 615
Xianghong Sun, Tom Plocher, and Weina Qu

Usability Evaluation of Children Edutainment Software 622
Danli Wang, Jie Li, and Guozhong Dai

Effect of Different Modal Feedback on Attention Recovery 631
Min Cheol Whang, H.J. Hyun, J.S. Lim, K.R. Park, Y.J. Cho, and J.S. Park

Do We Talk Differently: Cross Culture Study on Conference Call 637
Xingrong Xiao, Chen Zhao, and Shaoke Zhang

The Mobile Phone's Optimal Vibration Frequency in Mobile Environments 646
Jinho Yim, Rohae Myung, and Byongjun Lee

A Comparative Study of Mid-market IT Customers in China and U.S. 653
Yi Ren Yuan and Thomas Hogaboam

Author Index 659

Panel Discussion: Global Innovative Design for Social Change

Nuray Aykin[1], Apala Lahiri Chavan[2], Susan M. Dray[3], and Girish Prabhu[4]

[1] The New School, 55 West 13th Street, New York, NY, USA
aykinn@newschool.edu
[2] Human Factors International – Asia, Chemtex House, 4th Floor
Hiranandani Gardens, Powai, Mumbai 400 072, India
apala@humanfactors.com
[3] Dray and Associates, Inc.
Minneapolis, MN, USA
susan.dray@dray.com
[4] Asia PDC – India, Emerging Markets Platform Group, Intel Corporation
136 Airport Road, Bangalore 560078, India
girish.prabhu@intel.com

Abstract. As designers, we are solution seekers and innovators. It is in our core to find the best method or design to meet the needs of the customer, or create a great intuitive product that brings the most revenue. However, most of the work is concentrated on designing products for the people in the developed countries who could afford luxuries like the iPod and alike. There is a great shift now towards reaching beyond borders, especially designing for the people at the bottom of the pyramid. In this panel, we will concentrate on two areas that the design can play a significant role in advancement of societies: (1) Design for improving socio-economic structure such education, health, food and shelter, (2) Design for creating commercially viable products that can create sustainable businesses. Our panelists will share their experiences on how we, as designers, can make a difference in the way people live their lives.

Keywords: Bottom of the Pyramid, innovation, design, social change, social advancement.

1 Introduction by Nuray Aykin

The struggles of the world's poorest populations have, until recently, only been on the agendas of a few Non-Government Organizations (NGOs), aid agencies, national governments, non-profits, and individuals. However, a recent monumental change in thinking asks the world to view those living at the Bottom of the Pyramid [1], not as passive victims, but as active consumers, capable of identifying opportunities and creating innovative solutions. Innovation, as the core emphasis of this movement, focuses on reinventing business processes, life practices, news ways of solving problems, and building entirely new markets that meet untapped customer needs. This way of thinking calls for people to become active participants in the movement towards improving their own lives and well-being in addition to advancing in the economic pyramid.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 3–9, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Innovation leads to economic improvement, therefore to social change. Focusing on reducing the causes for diseases, improving socioeconomic life and supporting sustainable environments are becoming a strong mission for many designers. Designers without Borders, Massive Change are great examples of such kind. By concentrating on creating tangible outcomes would make this movement even stronger and would allow researchers and designers to more efficiently evaluate and measure the impact, and develop understanding of the complex relationship between people and their environment.

There are two areas that innovative design can play a crucial role:

1. Improving socioeconomic structure such as education, health and infrastructure. The design can aid solve water quality and supply problems, can bring a solution to agricultural issues, or can create ways to sustain the environment while providing the basic necessities to move the people into better conditions of living.
2. Creating commercially viable products to create sustainable economies. There are hundreds of examples of creative solutions that changed the lives of millions, including ultra cheap phones in India by Nokia, AMUL milk in India becoming a world brand creating millions of jobs, Unilever's project Shakti, affordable solar power units in Honduras bu Soluz Inc.

In order to succeed in this area, it is crucial to have strong ties with the state and local governments, NGOs, foundations and private organizations, and the people who are impacted by the social structure and are willing to participate in a long journey to move up on the economic ladder. As a designer, we need to educate ourselves to understand how multiple disciplines interact to create solutions to people's needs. We need to understand what the role of innovative design in social change especially related to environment, economic and health related issues.

In this panel, the participants will share their experiences and the lessons learned during their field studies.

2 Apala Lahiri Chavan's Statement

Even though Professor CK Prahalad pioneered the notion of companies targeting the lowest rungs of the market way back in the mid 1990s, it was after his book The Fortune at the Bottom of the Pyramid was published about a year back, that the concept gained increasing momentum.

His key argument: the so-called Bottom of the Pyramid (BOP) with an estimated 4 billion people who live on less than US$1,500 per annum, is a major market opportunity.

Not surprisingly, a number of Indian and overseas companies have tried to adopt this innovative business model. Hindustan Lever Ltd (HLL) has increased its market thrust behind Project Shakti, the low cost distribution model, which it already had, to target a wider base.

ICICI Bank has led a number of new initiatives to provide a host of banking services at affordable costs to the poor and lower middle class, including setting up a network of around 8,000 self-help groups.

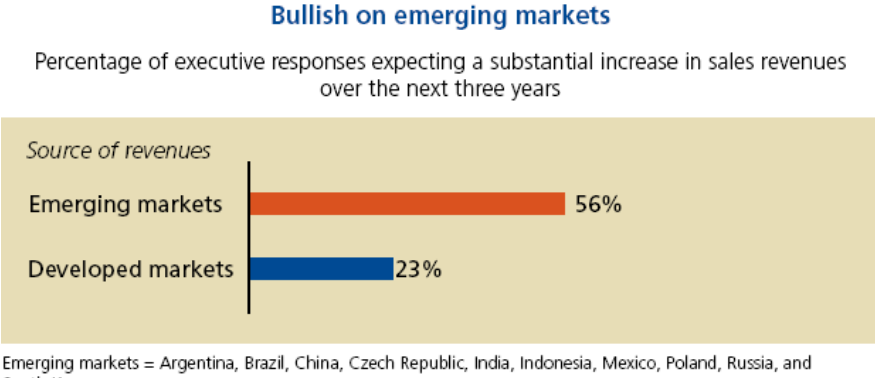
ITC is banking a lot on its eChoupal system for targeting rural farmer-entrepreneurs aimed at improving the agricultural supply chain, cutting supply costs, upgrading the information base for farmers and doing e-commerce.

At last count, the initiative was estimated to target over 3 million farmers through 5,200 installations covering 31,000 villages across six states.

Business Standard/ New Delhi October 07, 2005

Developing countries (India, China, Brazil etc.) are well known for low-cost manufacturing and providing customer support. The same countries are now considered by global companies as emerging markets for selling their products and services. Targeting the emerging markets is looked at as a way to make the competition and saturation in the developed markets irrelevant. Two major factors provide an opportunity for growth in emerging markets. One is the large size of this market. India and China together had up to 457 million households in 2002. If urban and rural consumers are put together, India alone has 100 million households.[1]

The other factor is, though out of these 100 million households in India the urban population is just 24 million, and average annual income is less than $6,000 per person, the purchasing power of these people is relatively very high. The behavior of this large group of consumers, by Western standards, is unusually youthful, demanding, open-minded, and adventurous.



Emerging markets = Argentina, Brazil, China, Czech Republic, India, Indonesia, Mexico, Poland, Russia, and South Korea

Source: One or more member firms of Deloitte Touche Tohmatsu

To penetrate this market, companies will have to go beyond mere adoption and localization of their products made for developed markets and take a radically new approach for designing, developing and deploying their offerings. A growing number of such companies now acknowledge that taking a radically different approach is the only choice in emerging markets as the consumers and the contexts in which their products will be used are totally different from the one in developed markets.

This attention by major corporate giants to the 'bottom of the pyramid' in the emerging markets have helped propel the aspirations of a group of people from the "top of the bottom of the pyramid" to leap forward to the next level of the pyramid or create a totally new level in the pyramid which never existed before.

[1] National Council of Applied Ergonomic Research, India: 2001-02 projection.

In order to understand this phenomenon, a joint project was initiated by the Institute of Design, Chicago and HFI, Mumbai. HFI has continued to work in this space till date.

The joint project aimed to observe daily lives in the homes of the top-tier people of the BOP in India and create design solutions that would improve these homes**.**

We attempted to understand the:

- Needs, Motivations, Aspirations and Attitudes
- Choke points, Pressure points and Pleasure points

And hence profile the user population from the point of view of design solutions that would improve the space that was 'home'.

Since the completion of the joint project, HFI has continued with the *"the leap forward"* aimed to conduct deep dive observation of a specific segment of the 'top of the BOP' (in this case, the potter community who live and work in Dharavi, the largest slum area in Mumbai) and provide innovative "out of the box" solutions catering to the changing needs and attitudes of this particular segment.

The "leap forward" questioned the basic needs, identifying latent and unarticulated ones which are emerging gradually in a mobile and customizable world. This project also tried to look into various aspects of technology and how it could be humanized keeping in mind, the future orientation of the target users.

In this panel, we will cover:

- Methods used for the study including description of the families, homes and their daily activities, perceptual mapping of their activities, our disposable camera study, tour of the house, future oriented discussions, and our debriefing with the families.
- Key characteristics describing "Top of BOP" such as restlessness in wanting to climb higher, future orientation, lifestyle, status, inspiration, opportunity seeking, optimizing the use of limited resources
- Key attributes that were articulated by participants as being important dimensions towards success (acceptance/recognition, accessibility, adaptability, alternatives, aspirations, betterment, community, compromise, constraints, convenience, family, future orientation, stability, opportunities, optimization, permanency, security, survival, status and lifestyle)
- Impact of innovation in terms of facilitating upward move in status (Status), supporting transitory lifestyle (Mobility), supporting low cost value additions (upgradeability), supporting multiple use and sharing (flexibility), facilitating community engagement (Collaboration)

3 Susan Dray's Statement

Many companies and organizations want to create innovative designs that can have a positive social impact. Indeed, there are literally thousands of examples of well-intentioned people working to provide access, technology or services to currently underserved populations around the world in the hopes that, by providing these things, they will help to create positive social change by providing economic or social benefits to the ultimate users/recipients. With zeal and money, these organizations

have high ideals and wonderful goals. However, many of these efforts fail to deliver on the promise.

Why is this? I believe that it is because, however clever or ingenious they may be, they have failed to take the entire context (economic, political, physical, infrastructure, organizational, social, familial, educational, etc.) into account sufficiently. The key to truly innovative design for social change is to first deeply understand the context, in all its myriad forms. This should be obvious to the user-centered design community, although sometimes even we are too narrow in our own definitions of "context" and limit our own explorations and research to understanding individuals or small groups (e.g., families) without taking these other aspects into account.

One example of a break-through service which has succeeded, by their own admission, *because* they took time to deeply understand the context, is Cell-Life, an NGO in South Africa. (For more information, check out http://www.celllife.org/) Cell-Life describes itself as "a pioneering initiative that provides effective technology-based solutions for the management of HIV/AIDS" in South Africa. Specifically, Cell-Life has developed an infrastructure for supplying anti-retroviral (ARV) drugs to fight HIV/AIDS, for tracking side effects (critical in determining future doses), and for monitoring drug compliance by patients by providing HIV/AIDS home care workers in rural and urban areas with cell phones. These home health workers visit patients and use a menu-driven mobile phone interface enter data about the patient's reactions to the most recent dose of ARVs, including side effects and symptoms, as well as drug adherence, and send this data using short message service (SMS) to a central data base where it can be tracked by a doctor and a pharmacist. This provides direct information from those closest to the patient to medical staff, often located at a distance. This represents a significant breakthrough in the number of HIV/Aids patients who can effectively receive ARVs, even though they live in rural areas.

This may seem like an "obvious" solution in retrospect, especially since South Africa has one of the most extensive cell phone coverage in the world's.extremely widespread cell phone coverage. Dr. Ulrike Rivett, Cell-Life's founder, estimates that 99% of South Africa has cell phone coverage. However, other cell-phone-based systems have not been successful. For instance, simply using SMS to send messages to HIV/AIDS patients to take their medicines has been tried and has not been so successful. Why, then, is Cell-Life's approach such a success?

According to Dr. Rivett, Cell-Life has been successful specifically because they deeply studied the entire context of the HIV/AIDS problem in South Africa before designing a solution [2]. They quickly realized that there were complex systemic challenges, in legal, political, and medical realms, which had to be addressed for any new system to succeed. Specifically, South Africa's constitution mandates access to health care for all. So far so good. However, in a country where many people live miles from paved roads, "access" can be a significant barrier. In addition, all medications must be dispensed by licensed pharmacists, who are in short supply especially in rural areas. Plus, AVR drugs also are not "standard" medications: They require must be refrigerated and the doses vary depending on the patient's reaction to previous doses and their current symptoms. Therefore, to dispense future doses, the pharmacist needs hands-on information about the patient to determine the correct dose.

They also must know for certain that the patient has actually taken the drug as prescribed. Without this type of information, they cannot dispense the medications [3].

Enter the Cell-Life team. After spending significant time understanding this context, Dr. Rivett and her team developed the Cell-Life system to give to pharmacists the information needed to prescribe and to doctors, the information needed for long-term treatment. They understood the serious obstacles to data capture in rural Africa, caused by inadequate infrastructure (intermittent electricity, poor roads, low bandwidth, etc.), low computer literacy and the need for training, and of course, cost. In addition, they understood the needs of stakeholders from a variety of communities, including medical and healthcare professionals, home health workers, patients, government officials, and technologists.

The resulting system was first piloted successfully, and has been adopted by the government of the Western Cape where it is being rolled out extensively. This has resulted in significantly more patients getting effective ARV treatment. The team has received accolades from many places, and news coverage by the BBC and others [4]. But perhaps the most telling is that the HIV/AIDS home health care workers have become among the biggest advocates for the system, for it has not only made life better for their patients, but it has also empowered them to play a bigger and more satisfying role in this care.

4 Girish Prabhu's Statement

According to Wikipedia, social change is change in the nature, the social institutions, the social behavior or the social relations of a society, community of people, or other social structures. The term covers concepts as broad as paradigm shift, to narrow changes such as a particular cause within local government. Though research in sociology suggests social change is created by various agents such as direct action, protesting, advocacy, community organizing, revolution, and political activism, the primary agent of social change is technological advancement. The wide adoption of a new technology leads to imbalance in the economic relationship between economic agents, leads to changes in the social balance of power, therefore leading to social change.

I believe design innovation plays a major role in social change along with technology. It is a well known fact that technology adoption does not happen unless it is designed to meet user needs. In emerging markets, especially for BOP (Bottom of Pyramid) and MOP (Middle of Pyramid), design has much more significance as these needs are at a confluence of social, cultural and economical aspects of people lives. Various dissonances in each of these vectors can lead to a slower pace in social change. I define dissonance as the gap between the intended usage models of the technology and the actual usage model. My hypothesis is that by reducing these dissonances through design innovation, technology can be utilized to create social change at a faster pace.

We explored the value of design in technology adoption for social change in a recent project. The aim of this study was to understand the needs of current technology owners (PC owners) for the development new ICT platform for middle tier and top of the bottom tier population of India. The primary task was to find out what is it that makes people from emerging economy countries not only desire to

buy/own PC/Technology but also use it to make a difference in their lives. For example, people may buy/own a product/service as a status symbol but may not use it in its intended usage model. This was termed as Technology Dissonance. The study revealed a broad set of dissonances:

- Dissonance due to perception of technology: Perception of the PC is found to be one of the most important observed dissonances since it contributes directly to the mental model and technology adoption. The factors that feed into this are the issues of fragility, complexity, technology fear and the fragmented form of PC itself which creates operational problems as well.
- Design Dissonance: These included design issues that make the current PC platform a misfit for the emerging markets. For example we found that PC is not designed for ease of use like an appliance, suffers from lack of local language languages support, is not designed for group usage (which is extremely prevalent in India), provides low flexibility and does not sufficiently address the needs of mobility and connectedness.
- Usage Dissonances: The issues in this category speak of a varying pattern of usage among various household segments and work domains. The reason for the changing nature of usage can be traced to unique socio-cultural attributes of these sections. A closer look at the priority of usage indicates that various socio-economic segments of households and small businesses put different emphasis on the broad level needs. This indicates a high diversity in the functionalities that are probably needed in addressing the user needs.
- Cost and ROI Dissonance: The current product has high acquisition costs with perceived, frequent and substantial costs of software and hardware up gradation. Apart from the tangible costs it creates a fear among the users of emotional costs involved in future up gradation.
- Eco-System Dissonance: Ecosystem related dissonances arise from the factors in the surrounding environments that include lack of service & support from the PC sellers, poor power and internet infrastructure and lack of information among the general consumer about the available PC products and services in the market.

This research suggests that users at the BOP and MOP expect more from technology apart from design and usage congruence for technology to be adopted. Factors such as clear perception of the value, fit of the technology in the surrounding environment, and also business value proposition play a major role in adapting technology for change in their lives. And hence design innovation has a major role to play.

References

[1] Prahalad, C.K.: The fortune at the bottom of the pyramid: eradicating poverty through profits. New Jersey, The Wharton Press Paperback Series (2006)
[2] Rivett, U.: Personal communication (2007)
[3] Anand, S., Rivett, U.: ICT in the management of HIV treatment: Cell-Life: A South African Solution. The Journal for Convergence, vol. 6(3). Available for download at: http://www.celllife.org/
[4] Lindow, M.: How SMS Could Save Your Life. Wired News (November 4, 2004) Available at: http://www.wired.com/news/medtech/1,65585-0.html

Enabling User Centered Design Processes in Open Source Communities

Mads Bødker, Lene Nielsen, and Rikke N. Orngreen

Center for Applied ICT, Copenhagen Business School,
Howitzvej 60, 2000 Frederiksberg, Denmark
Mb.caict@cbs.dk, ln.caict@cbs.dk, rno.caict@cbs.dk

Abstract. Drawing on tenets from action research, this paper presents a yearlong intervention designed to facilitate knowledge of actual users and use in an Open Source Software (OSS) development community. Results from the interventions are presented and the influence of central characteristics of the OSS community and its communication is discussed. Initial findings show that the ideology and praxis based approach of the OSS community, as well as their primary media of communication, present a challenge to the introduction of end-user issues.

Keywords: Open Source, usability, developers, community, learning, action research.

1 Introduction

This study reports from a project that aimed at introducing usability and user awareness into an Open Source Software (OSS) developer community. In the field of human computer interaction (HCI) a variety of methods exist that focus on user centered design. A wide range of methods exists that investigates user needs and context as well as methods that involve users directly in the development process (participatory design) [7], but these have not yet been widely applied in open source development. This potentially leaves a gap between the developer-users, those who extend and innovate on the OSS, and end-users, those who will end up actually using systems for their intended purpose. The question for the following is, if user-centered design thinking is something that can enter into the Open Source development environment?

Existing usability studies in connection to OSS development focus on how users report bugs and wishes for new systems features as well as how the development community reacts towards these reports and wishes [2, 10]. Other studies on user involvement in design processes often focus on user driven innovation, where involving users in the development process using a variety of methods are claimed to bring about new and innovative designs [6]. In particular of interest to the open source community are the so called Innovative communities [5]. As such, the OSS development community qualifies internally as an innovative community – there is a strong element of competition and innovation in OSS development. However, the user

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 10–18, 2007.
© Springer-Verlag Berlin Heidelberg 2007

in the case of OSS innovative user communities are the before mentioned developer-users, not actual end-users of technology. In this way, innovation in the OSS community does not extend across the boundary into the context of those users who use the systems on a daily basis for mundane and leisurely task. Correspondingly, knowledge of end-user situations can not be seen to play any significant role in the development of the system and potentially not arriving at end-user friendly products can pose a serious threat to popularity and adoption of the system.

This paper introduces a yearlong intervention project carried out through most of 2006 with members of the Open Source community behind the Content Management System TYPO3.

2 TYPO3

TYPO3 is a widely used small to midsize enterprise class content management system (CMS) under an Open Source license. TYPO3 has been publicly available for 5 years, and it has currently approximately 320 active contributors.

The TYPO3 community has never signed up for a formal membership. Rather, it consists of people who join the TYPO3 mailing lists, newsgroups and more formal groups, for example the R&D group, the Core development group and so on. The members are a highly diverse bunch: some are highly skilled programmers who participate with an interest in developing system extensions; others are interested users who use the mailing lists to put questions about use. The community is organized in several subgroups and communication takes place in discussion lists as well as in occasional physical sub-group meetings. The discussions seen at the TYPO3 community (see the typo3.org website) are generally oriented towards the implementation of extensions to the system or which bugs should be fixed.

On the typo3.org website the R&D group, in its own words, state that their aim is developing a system which is complex and yet usable so as to support business CMS solutions. At one point, the team had chosen to address the issue of usability in their coming TYPO3 versions, but realized that the “code now, humans later” [15] focus of the developer community made it difficult for them to attract the knowledge needed. This problem made the R&D group approach the authors of this paper in order to initiate a process of introducing usability awareness to part of the community.

3 The Method

The research presented in this paper is based on principles and ideas from action research. The aim of action research is to create change by improving a specific case, in a specific period of time, at a specific location [17].

The action research label accounts for a number of different attitudes towards research process and methodology. While the change-oriented contention of action research is central, there is a great variety and no methodological canon to be followed. [1]. Our central aim in adopting an action research approach to the project was that we did not only want to study the state of the art in the developer community, but throughout the study maintained a therapeutic stance, wanting to change the orientation of the developer community.

Our project fell in two distinctively different phases that we have termed the Ambassador Project and the Learning Project. Both phases took place in the context of a HCI discussion list we had set up to officially indicate that an initiative directed at improving TYPO3 usability was in progress.

In the first intervention, the Ambassador Project, we as researchers would have to get an understanding of the users and use situations. Then we would ask the ambassador participants to investigate their users and share their knowledge with the other developers [12]. These were to be used as a basis for personas descriptions intended at the distributed development process.

In the second intervention, the Learning Project, we introduced a set of heuristics in order to provide the participants on the mailing list with a common vocabulary for usability, supposing that having some form of contextually relevant knowledge on usability, equally available to all developers, would set some form of reflection upon end-user issues in motion.

4 First Intervention: The Ambassador Project

As with most other OSS projects, the TYPO3 development structure consists of developers who carry out programming of the core TYPO3 system (the stand-alone system) and scores of developers in the various user-groups who use the TYPO3 source-code to program individual business solutions. In practice, developers may take on both roles – doing coding for the sake of the TYPO3 system itself and suggesting new features found usable during individual projects, suggesting them for implementation in TYPO3.

From following discussions on the TYPO3 HCI-list, it quickly became apparent that there was no explicit and common knowledge in the community of whom the actual end-users are. As part of the process of creating awareness of the users, we conducted a pre-study of the use of the TYPO3 CMS in two organizations differing in size, complexity of the TYPO3 system implemented, and the end-users' possibilities for IT support. Four interviews and three videotaped observations were made. Talking to end-users and seeing them use the system provided insights about work, work situations, and attitudes. Attitudes originated in computer skills with end-users being either comfortable with computers thus putting demands to the system or uncomfortable with computers, but pleased with the system as long as fixed procedures were followed.

Table 1 below shows an excerpt of a description of an actual end-user, based on a process of meaning condensation [2] from our interview with and observation of an end-user in a large public organization using TYPO3. Sarah, who is a highly skilled technological user, is focused on how the system fits her working processes and needs.

Descriptions similar to this, derived from our observations, were used to draft HCI ambassadors. Our assumption was that poignant examples of real, "lived" experience of TYPO 3 use could attract developers with particular interests in usability and end-users to act as ambassadors on the list. Coming from the same "programmer culture" as the other developers on the list, we assumed that these ambassadors would be better equipped to disseminate their interest to the wider community.

Sarah is a legal advisor, employed in a large Danish public organization and uses TYPO3 everyday. Her attitude towards the system is that it has not been designed to accommodate for her way of working and the tasks she has to perform. She has to scroll far too much among the documents and she has many ideas of how to improve the system.
Sarah gives an example of one of her tasks carrying out statistic logs of legal decisions, where she states that "this is a real nerd-calculator, top-nerd! ... so I can't use this for anything" and continue to show how she performs the calculation manually. She has tried to speak to the IT department about improving their adapted version of TYPO3, but feels that they don't understand her. As she explains: "They speak Chinese and we speak Danish and there seem to be no dictionary"
When she started using the system she spent some time on an in-house course, but she has mainly learnt the system by using it. This also makes her create her own shortcuts and "this is why I have developed my own ways of working with the system", she says when pointing to some of the obstacles of the systems features. She is a person that others contact when they are stuck. Recently she spent two days training a newbie to TYPO3, as Sarah has been offered a new job. This was quite a frustrating experience and as she says in her new job: "I'm never going to play with TYPO3 again"

Fig. 1. Except from a description of an actual TYPO3 user

On the HCI list, we asked the developers to consider what they knew about their end-users, and to submit written descriptions of actual end-users they had met. They immediately perceived the request for user descriptions as a request for descriptions of abstract user-types, which they denoted "personas" in their discussions. Taking a solution-oriented approach the developers used these personas to describe solutions for the system. Later they were asked to interview a selection of users and four e-mail interviews were carried out. These interviews showed that most end-users were content with the system, but they also exposed a huge variation in the use of the CMS. Either it was used by novice users with a very limited set of functions on a less frequent basis, or by users with high computer skills, using a wide range of functionalities on a daily basis. This supports the observations made earlier, but the interviews were too few to be of any actual value. While the ambassadors were well versed in communicating on the HCI-list, they lacked knowledge on usability concepts and aim, and even if they found it to be important, it never became clear to them what the aim of the project was and no more data came out of it. This made us close down the project to continue along another line of intervention.

5 Second Intervention: The Learning Project

The correspondence on the HCI-list exposes a frequent inability to cope with engagement in end-user issues other than by implementing rapid solutions to clearly specified problems. To the developers, engaging in users is seen as a solution oriented problem, since end-users are perceived as solution finding actors. Taking over where the ambassador project ended, we decided to use the HCI list more systematically to

facilitate a learning process amongst developers enthusiastic about HCI. Primary in this process was the use of the mailing list and the TYPO3 wiki to assist the group in creating a shared vocabulary and provide clear examples of usability thinking that serve as guidelines for deliberating and subsequently solving problems.

The HCI-list has developed steadily since May 2006 and it still features a lot of discussions about solutions with a noticeable exception being a discussion taking place in late September 2006. A thread started by TYPO3 founder Kasper Skårhøj whose re-reading of an article [1] instigated the asking of more critical questions about end-users. The thread can be distinguished from others in the HCI-forum since it sought to determine which solutions are better considering end-users and the motivation of developers to solve end-user problems. While posts about specific solutions get more attention in terms of replies, this discussion occasioned a rather extended dialogue consisting of 24 posts from 11 different posters. However, the problem seems to be that discussions lapse towards either specific problems (e.g. labeling of functionality), towards paradigmatic observations of a very general nature (e.g. are we "dumbing the system down" or are we making it smarter?), or towards ethical paradoxes inherent in open source development (e.g. why care about users at all when you do things for free?).

Since our analysis of the discussions indicated a generally poor understanding of the concept of usability and as the community seems mainly to concentrate on technical solutions, we tentatively introduced 10 heuristics (see wiki.typo3.org/index.php/Heuristics) derived from [11], [14], [9], [16] and chosen amongst the many principles for design introduced by the authors to reflect typical problem areas in the specific TYPO3 CMS domain. With a detailed description of the heuristics posted on the mailing list and available to the discussants on the TYPO3 wiki, we were hoping to facilitate a shared vocabulary for the developers, a common place of reference enabling a process where problems with the TYPO3 interface were no longer seen as highly specific, but as indicative of more principal problem fields and hence applicable to heuristic analysis. A shared vocabulary and some basic knowledge of more abstracted concepts in understanding how users interact with systems, so we expected, would raise the bar for the discussions on the list and potentially make the problems discovered eligible for shared solutions rather than the unsystematic and narrow focus of solving "one-off" particular problems.

6 Analysis

Introducing concepts of usability and, more broadly, an understanding of- and empathy with users into the OSS development community, proved to be a challenging undertaking. In the following we will analyze and evaluate how the community of TYPO3 developers interacted and how our intervention was used in the community.

Since the development of TYPO3 is Open Source, we find it necessary to look at a network of OSS discourse, ideology, and praxis to see how these can be said to conflict with our style of intervention and the possibilities for change and learning. Further, we will assess the medium wherein learning and communication was facilitated.

6.1 OSS Discourse, Ideology and Practice

Even if the project did succeed in setting up an active list for HCI interested developers (the HCI list), a common objection on the mailing list was "why should we develop for "users" (meaning here end-users) since what we do is essentially for free and since we do it simply because we like programming, why should we care how or how well "regular" users use the things we build? For instance, HCI-list postings along the line of *"a core developer has no responsibility above whatever his personal motivation may be"* (Oct. 13. 2006) or *"why is it not naturally for everyone to scratch ones itch?"* (Oct. 12. 2006). This line of reasoning is reminiscence of what we could call classic OSS discourse or ideology. Using a concept of ideology as a normative structure that tacitly and seemingly a-historically allow us to think and believe in specific ways, OSS ideology seems to rely strongly on classical democratic tropes of sharing and equal relationships between peers. Sharing and transparency is key terminology used to describe the nature of a working OSS development community [8].

Yet since OSS is by definition developed "con amore" and with no direct economic incentives, there is no perceived obligation to actually "care for the itches of others" - to have any kind of empathy for those outside the loosely coupled group of developers who share knowledge, skills, values and vocabulary. As Eric Raymond, who co-coined the term Open Source states in his seminal book "The Cathedral and the Bazaar": "Every good work of software starts by scratching a developer's personal itch" [13]: p. 23.

Therefore sharing and transparency are attributes that are at work within the community of developers themselves, not something that has any relation to end-users. In short, we can say that the Open Source encouragement structure and the non-hierarchical community arrangement and the strong sense of emotional belonging that the community commands tends to preclude the possibility of seeing beyond their own motivations. Thus while the originating ideology of Free and Open Source software development seems to hinge on an altruistic, purified democratizing effort and a cleansing of capitalist incentives in the development of technology for the coming Information Society, there is a marked element of neo-liberalist thought that disqualifies the perception of ones work as a service to end-users, those who are not themselves part of the development community. Hence, the community ideology itself presents a challenge to the introduction of user-centered thinking – there is simply no obvious incentive.

6.2 Praxis

Another aspect that posed a severe challenge to our intervention was that while our project sought to activate the developers reflexivity of other peoples use of their product in order to facilitate more user-centered design, OSS culture is, as De Joode has pointed out, a culture of *doing* not of *deliberating*, a specific communal trait that we will also discuss in the next section. OSS development is a culture of proving ones

worth in practice, not of using abstract ideas to guide ones practice. OSS development is a zero-sum game where the provably best piece of code is adopted into the system while the less functional ones are abandoned [18]. Compared to that, user-centered design, even when taking up empirical modes of inquiry is quite different when it comes to proof or proving. While user centered design experts can make educated assertions about how interaction will take place, and hence derive a set of heuristics from these assertions, a user-centered product is not static in the same sense that a provably more efficient algorithm is. This difference in the culture of proving and the praxis derived from it could be seen as one of the reasons for the failure of our initiative to introduce principles of usability to the community.

6.3 Communication – The Medium

The primary coordination and communication tools used in OS development communities are e-mails, mailing lists, forums, and other forms of digital networked communication tools [8]. So too, in the TYPO3 HCI community where the primary communication took place on the HCI mailing list. While the TYPO3 development structure does have a certain hierarchical organization, using an appointed association to take decisions on official TYPO3 releases and certification of commercial TYPO3 agencies, no hierarchical organization existed to enforce decisions or to evaluate the outcome of the ongoing discussions on the mailing list. As such, the list provides a space for a kind of *ideal speech situations* [4] where no external criteria are used to evaluate the rationality of communication. However being principally unperturbed by external power also disenabled decision making, as the discussion on the mailing list itself was not able to make decisions that cut across group boundaries, for example across those who are in favor of making the interface less complex and those who favor extensive end-user programming. Acting mainly as a pure discussion forum, no efforts were put into enabling e.g. consensus decision making such as it is carried out in other open-standards and open source organizations (e.g. W3C, see [19]). As we have seen, this resulted in the list being used to primarily share specific solutions to specific design problems, and to share immediate problems, which could potentially be solved by other participants on the list.

While the power-free communication of a mailing list could be said to be a decent and indeed moral procedure highly consistent with the OSS community ideology, it did not, in our case, meet the criteria of actually pushing the innovation of TYPO3 usability forward. Since it was mainly used to assist in the solving of individual users problems, and hence enabling them to better “scratch their personal itches”, our analysis shows us that we did not adequately assess the problems inherent in using the mailing list to facilitate usability learning and subsequent innovation. Innovation here should be understood as the introduction of a considerably more user-friendly interface in coming official TYPO3 releases. One way to solve this problem of on-line, distributed decision making could have been to facilitate consensus based decision making, entailing for instance that propositions should be clearly marked as such, thus being eligible for assessment and consensual verdict, or by appointing a managing committee to enforce some form of conceptual integrity [3] – making sure that the hundreds of ideas that surface on the list correspond to one overarching goal.

7 Conclusion

OSS development, while proceeding directly from the practical and technical tenets of software engineering, has since the early days of hacking departed from traditional management principles such as those presented by Brooks in his book on the Mythical Man Month [3]. Where Brook's assertion was that adding more programmers to a delayed project slowed down development, Eric Raymond argues that the OSS strategy of distributed development has other virtues. Certainly, as he argues, "given enough eyeballs, all bugs are shallow", indicating that distributed bug elimination, loosely coordinated via the Internet, is indeed an efficient strategy. The question is if the same anarcho-libertarian OSS tactics are as efficient when it comes to designing for "real" end-users?

A number of things came of our intervention. First of all, the fact that issues pertaining to usability are explored and discussed at all. Since the personal motives in OSS development are primary to all other incentives, this can be seen as an innovation in the TYPO3 community. Secondly, the participants on the list began as of late November 2006 to work on a comprehensive survey to assess how users experienced the usability of TYPO3. The real innovative aspect of this survey was that it was highly attentive to the fact that there were indeed many kinds of users, and that, in order to communicate and gain insight into a multitude of users, different kinds of language should be used and different kinds of questions were needed. Before our intervention we found the development team framing their discussion of users as a discussion of the conflict between designing for "dumb end-users" or for highly skilled "administrator-users". The survey and the associated discussion on the list suggest that our intervention pushed some developers' attitude from antipathy towards empathy with end-users.

A pertinent lesson to be learned from our inquiry is that the developer segment in OSS is not particularly disposed to concern themselves with phenomena outside of the community. Rather than addressing the developer segment, it might be conceivably more sensible to address the actual user segment. Enabling the organizations and institutions that make use of OSS software to understand their employees and enabling them to specify the right requirements for their system, might be a more efficient way to avoid developer-centric systems that perform poorly in terms of real-world usability.

References

1. Baskerville, R., Wood-Harper, A.T.: A Taxonomy of Action Research Methods. PAP012-0.05 (1996)
2. Benson, C., Müller-Prove, M.: Mzourek, Jiri.: Professional Usability in Open Source Projects: GNOME, OpenOffice.org, NetBeans. CHI2004. ACM, Vienna, Austria (2004)
3. Brooks, F.P.: The Mythical Man-Month: Essays on Software Engineering. Addison-Wesley, Reading (1995)
4. Habermas, J.: The Theory of Communicative Action. Beacon Press, London (1981)
5. Hippel, E.V.: Democratizing Innovation. MIT Press under the Creative Commons Rights (cc) (2005)

6. Jeppesen, L.B.: User Toolkits for Innovation: Consumers Support Each Other. Journal of Production Innovation Management 22, 347–362 (2005)
7. Kensing, F.: Methods and Practices in Participatory Design. ITU Press, Copenhagen (2003)
8. Ljungberg, J.: Open Source Movements as a Model for Organising. European Journal of Information Systems 9, 208–216 (2000)
9. Moore, P., Fitz, C.: Gestalt theory and instructional design. Journal of Technical Writing and Communication 23(2), 137–157 (1993)
10. Nichols, D.M.: Twidale, M.B.: The Usability of Open Source Software. First Monday, vol. 8, 1 (2003)
11. Nielsen, J., Molich, R.: Heuristic Evaluation of User Interfaces (1990)
12. Nielsen, L., Orngreen, R., Nielsen, J.: Engagement in Users - a new approach to open source development. International Design for Engagement, Oslo (2006)
13. Raymond, E.S.: The Cathedral and the Bazaar. Musings on Linux and Open Source by an Accidental Revolutionary. Revised edn. O'Reilly & Associates, Inc., Sebastopol, CA (2001)
14. Shneiderman, B., Plaisant, C.: Designing the User Interface. Pearson, Harlow (2005)
15. Skårhøj, K.: TYPO3 - presentation to HCI students. E-business, IT University, Copenhagen (2005)
16. Tognazzini, B.: First Principles of Interaction Design. AskTog http://www.asktog.com/basics/firstPrinciples.html (2003)
17. Toulmin, S., Gustavsen, B.: Beyond Theory. Dialogues on work and innovation, vol. 2., John Benjamins, Amsterdam (1996)
18. Van Wendel De Joode, R.: Understading Open Source Communities - an organizational perspective. Technische Universiteit Delft, Delft (2005)
19. W3: http://www.w3.org/2005/10/Process-20051014/policies.html#Consensus

A Dramatic Day in the Life of a Shared Indian Mobile Phone

Apala Lahiri Chavan

Vice President - Asia
Human Factors International
Chemtex House, 4th Floor
Hiranandani Gardens, Powai
Mumbai 400 072
apala@humanfactors.com

Abstract. The paper explores the area of culture strain and how it affects the usage and hence the design of products and services. In this era of globalisation, it is increasingly important to create a tool kit of methods and techniques that will address cross cultural use of a product. This is particularly important in cases where the product is designed in and for a particular kind of culture and then it is 'exported' for use in widely different cultures. Till date, it has been common to 'localise' such a product by looking at the dominant cultural characteristics of the culture where the product is being exported for use. This paper takes the view that it is equally important to look at the culture (where the product is being exported for use) not just as it is supposed to be but also as it is. The difference between the 'cultural ideal' and 'cultural practice' [1] does indeed provide some rich opportunity areas for value added design solutions.

1 Dramatic Conflict

Three dramatic scenarios…each different but at the same time united by a common thread. The common thread is the cell phone and the way it is used. Unlike the cell phone being an extension of ones individual identity, as is the case in much of the western world and in some small pockets within Asian countries, these scenarios are all about the cell phone playing a very conflicting role! A conflicting role because of one inherent attribute of the cell phone that contradicts a dominant cultural attribute of the user population.

And that attribute is that of the cell phone being an individual device, ideally meant for use by one person. While this works in the western world, where it was originally designed, this attribute is at odds with the largely collectivist culture of the Asian countries.

1.1 Scene 1

Seventeen year old Amar just came back home, rather sheepishly. Its midnight and his parents are sitting in the living room pretending to watch television. Pretending because they are actually sitting up waiting for Amar to return. They are not happy at

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 19–26, 2007.
© Springer-Verlag Berlin Heidelberg 2007

all that he has come back so late. Of course, that they are at least not talking about disowning him is because of that magical device, the cell phone!

Amar borrowed his father's cell phone when going to the party earlier this evening. In fact, that he was allowed to attend the late night party at all, was on condition that he would be available at all times on the cell.

Now that he is back …a little later than promised, he faces another major problem! He has to now return his father's cell phone.and the reason that is a problem is because he has made 12 phone calls to his girlfriend. They had a slight disagreement this evening and it took a dozen phone calls to bring things back on an even keel. Not to forget the nine text messages.

In the hurry to drop her back home after the party and rush back home since it was already past the Cinderella hour, Amar forgot to delete the calls and messages. The only hope is if his father forgets about the phone and asks for it in the morning. If he asks for the phone right now, Amar could be in deep trouble. If his father were to see the messages and all those calls to the same number, he would know about his relationship with Leena and that would be a major disasater!

1.2 Scene 2

Deepa and Saurabh have got married recently and moved to Mumbai just 2 months ago. he is a software engineer and she is a homemaker. She has heard so much about Mumbai that she is very fascinated by the city. She spends a lot of time exploring interesting home stores. The more she reads about Mumbai and all the bold and beautiful people who live here and sees pictures of their homes, the more she desires to create an unique home for Saurabh and herself.

The only hitch is that Saurabh does not quite see why she has to buy so many things for the house that seem unnecessary. He strongly believes that as long as the home is functional, they are done with setting up the home. All these frills and fancies seem rather extravagant to him.

Deepa spends her own money to create her fantasy home. She has saved money from the time she used to work in a school before her marriage. So, while she does not have to ask Saurabh for money for buying all the nice little frills for her home, she does often give a lower amount than what she really paid, when Saurabh asks her the price of some new artifact that she has just bought. She knows that he will throw a fit if he knew how much she really paid!

Deepa was reading the newspaper in the afternoon when she suddenly saw the large adverts for the 'red sale' at the upscale Bandhni home store. Today was the last day! Deepa decided that she had to go right away. She had a quick lunch and set out for Bandhni. She could not inform Saurabh since he had borrowed her cell phone today (his was at the Nokia shop for some minor repair job) when going for work.

She shopped to her hearts content even though she knew that she had overspent. It was fine, she told herself. Such opportunities came rarely and moreover, Saurabh would not know how much she spent. Thank heavens for her credit card! And suddenly she froze…

'Oh my god!' Deepa broke out in a sweat… her cell phone was with Saurabh and she had just used her credit card to buy all these wonderful home artefacts! 'Oh no!' she groaned… the moment she used her credit card, her bank sent an instant message

to her cell phone stating the details of the transaction, that is, the money spent and the place where the transaction happened. She loved this feature because it made her feel so secure. She knew that if anyone misused her card, she would know instantly! But right now, she would give anything to have not had this feature!

Saurabh would , by now, have all the details of the money she had spent and what was worse…if he looked at her text messages he would have a whole list of messages that gave details of all past transactions. And of course, since she often did not tell Saurabh the correct price when he asked, what he would see on her cell would be rather different from what she had been saying..

Deepa felt the ground below her feet sinking …and frantically wondered what she should do…

1.3 Scene 3

Kunjipur is a large village in the northern part of uttar Pradesh. Many families from this and nearby villages have their menfolk working in various countries of the Middle East. They earn a lot of money and try to convince themselves that that sort of makes up for their absence from the family.

The village is a very typical Indian village, with scanty infrastructure but a lot of spirit and entrepreneurship. Most houses don't have land line phones or cell phones or permanent 'power' connections. However, most of them do own television sets.

Raju lives in a large extended family in kunjipur. His father and fathers' two younger brothers have all gone to Sharjah to work as plumbers. So he lives with his mother, sister, younger brother, grandparents , two aunts and 5 cousins. They live in the ancestral home which has been recently extended , with the money sent by the men from Sharjah.

All of them miss their fathers /uncles/husbands/sons respectively. They don't have telephones in their house and therefore to speak with their fathers or uncles, they have to walk to the crowded village square and queue up for the one public phone booth their village has.

The entire family is very excited because they have just received these interesting 'cards' from the bank. Raju's father and his uncles would bring a lot of money with them when they returned home for their annual leave. However, they always felt scared carrying the cash with them. They also sent some money by using the unofficial 'havala' channel. He has heard from his father than even the 'havala' is not a very safe way to send money.

When they came home last month, the entire family went to the State bank of India branch in the district HQ that was 30 km from their village. It was a picnic for the whole family as his father and uncles spoke with the bank manager for a long time. His father, then explained to all of them that each family would soon receive a card. That card, would, magically be able to get them money from the state bank whenever they needed.

The cards had just arrived, one for his mother and one each for his aunts. They were all with him since he was the only literate member of the household. He read the letters that came with the cards and he knew that there was a number called a 'Pin number' that he needed to have before they could use the cards.

Raju felt so powerful holding the three cards in his hand. He felt that everything depended on him…he would be the one to get the money from the bank and give it to his mother and aunts. Perhaps they would give him a small amount as salary to do this task ? Or perhaps he should just take it himself and not even ask them?

As Raju stood thinking about this wonderful new position he had carved for himself, he was rudely shaken out of his reverie by the arrival of postman uncle. Postman uncle came on his bicycle with all the letters. Postman uncle was like a family member. Raju rushed in to get postman uncle the tea and snacks he always had when he came to deliver letters. Now that raju could write letters , postman uncle did not have to write his mothers and aunts letters for them.

What was very exciting though was that postman uncle was coming home nowadays with a cell phone! The government had started the 'daakiya aaya, mobile laaya'(the postman is coming, he is bringing the mobile phone) scheme for villages which had poor land line telephone infrastructure. So he came with his cell phone and everyone could make a call using his cell phone, for a fixed charge that differed per country.

Once postman uncle finished his snacks, he sat with raju and the rest of the family to make an important call to his father, then to each of his uncles. Postman uncle called his father and raju was the privileged person who got to speak first. His father said that he would now be letting him know the magic number that would enable them to use the card. He would also send the numbers for his uncles' cards. But he would not say it out on the phone. Instead he would send a text message with the Pin numbers right away on postman uncles cell phone. He then gave detailed instructions to raju about what to do with the numbers and the cards when raju went to the bank next week.

At the end of the call, raju waited anxiously for the message to appear on postman uncle's cell phone. And suddenly, there it was! Three numbers for the three families. Raju felt like a king now. He had three cards and three pin numbers. Everyone had to depend on him for getting the money from the bank. He grinned at the many possibilities…

Postman uncle bid them goodbye and carried on to the next house. He was very happy because now he would show all the neighbours all these three numbers that raju's father just sent on HIS cell phone! Yes…there would be so much admiration for him for being such an important person!

2 Collectivism Defined

Collectivism is defined as one of the primary dimensions we often use to measure how cultures differ. The primary dimensions as developed by Geert Hofstede [2] are:

Power Distance Index (PDI) focuses on the degree of equality, or inequality, between people in the country's society. A High Power Distance ranking indicates that inequalities of power and wealth have been allowed to grow within the society. These societies are more likely to follow a caste system that does not allow significant upward mobility of its citizens. A Low Power Distance ranking indicates the society de-emphasizes the differences between citizen's power and wealth. In these societies equality and opportunity for everyone is stressed.

Individualism (IDV) focuses on the degree the society reinforces individual or collective achievement and interpersonal relationships. A High Individualism ranking indicates that individuality and individual rights are paramount within the society. Individuals in these societies may tend to form a larger number of looser relationships. A Low Individualism ranking typifies societies of a more collectivist nature with close ties between individuals. These cultures reinforce extended families and collectives where everyone takes responsibility for fellow members of their group.

Masculinity (MAS) focuses on the degree the society reinforces, or does not reinforce, the traditional masculine work role model of male achievement, control, and power. A High Masculinity ranking indicates the country experiences a high degree of gender differentiation. In these cultures, males dominate a significant portion of the society and power structure, with females being controlled by male domination. A Low Masculinity ranking indicates the country has a low level of differentiation and discrimination between genders. In these cultures, females are treated equally to males in all aspects of the society.

Uncertainty Avoidance Index (UAI) focuses on the level of tolerance for uncertainty and ambiguity within the society - i.e. unstructured situations. A High Uncertainty Avoidance ranking indicates the country has a low tolerance for uncertainty and ambiguity. This creates a rule-oriented society that institutes laws, rules, regulations, and controls in order to reduce the amount of uncertainty. A Low Uncertainty Avoidance ranking indicates the country has less concern about ambiguity and uncertainty and has more tolerance for a variety of opinions. This is reflected in a society that is less rule-oriented, more readily accepts change, and takes more and greater risks.

Long-Term Orientation (LTO) focuses on the degree the society embraces, or does not embrace, long-term devotion to traditional, forward thinking values. High Long-Term Orientation ranking indicates the country prescribes to the values of long-term commitments and respect for tradition. This is thought to support a strong work ethic where long-term rewards are expected as a result of today's hard work. However, business may take longer to develop in this society, particularly for an "outsider". A Low Long-Term Orientation ranking indicates the country does not reinforce the concept of long-term, traditional orientation. In this culture, change can occur more rapidly as long-term traditions and commitments do not become impediments to change.

3 Culture Strain

So as one can see from the definition of individualism/collectivism, cultures where this dimension is strong are cultures where ‘sharing’ is a very important part of life. This implies that inherently people who belong to collectivist cultures share personal space and objects much more than those who belong to individualist cultures.

With the advent of the cell phone, has emerged the contradiction between a collectivist population using a device designed for an individualist culture. Interestingly, this should have made it very difficult for, say the Indian population, to

use the cell phone. However, the rapid penetration of cell phones bears testimony to the fact that the cell phone is certainly very popular with Indian users.

What then do the scenarios described in the beginning of this paper really mean? All the 3 scenarios were about people using the cell phone in an individualistic manner (to a greater or lesser degree) amidst a collectivist 'ecology'. This led to the friction and edgy situations experienced by the 'actors' who were part of the scenarios.

Does this mean that in spite of a culture having a certain orientation, people can behave in a manner that contradicts the dominant orientation? The answer seems to be a resounding 'yes'. Cultures are not static entities and therefore they change and often over a period of time morph into an entity that is different from what it was a generation ago. Dr. Genevieve Bell, anthropologist at Intel believes that , in fact, the places where the tensions are strongest between cultural ideals and cultural practice are the most interesting. They're also often places where technologies are very successful (1).

We define this as culture strain, where the gap between what **ought to be** and what **is** creates dissonance and hence opportunities for design solutions.

Amar would have definitely liked an easy way to guard his privacy with regard to the calls he made and messages he sent using a shared cell phone, Deepa would have loved to guard her privacy with regard to the purchases she made and the resultant messages she received, when her cell phone was being shared and raju's family would be better off if they were guarded against raju's temptation to misuse the Pin numbers that he possessed.

4 Some Examples of 'Culturally' Dual Purpose Products

4.1 Cell Phone

In fact, the cell phone is a very good example of a device that has become very popular in both modes of usage , that is, 'mainstream culture' as well as 'counter culture' even in cultures which it was not designed for. The cell phone has become immensely popular in asia because it allows people to communicate and stay connected (very mainstream cultural attribute of this region) especially given the uneven quality and quantity of private and public land line telephone infrastructure. However, the cell phone's popularity in the region is also because it allows 'counter culture' behavior. Take the example of Asian women and the cell phone. The cell phone has allowed immense empowerment of women in the region by allowing women the freedom to converse and connect with anybody, anywhere…in private. This is very 'counter culture' behavior but became possible because the same device also met the needs of mainstream cultural requirements. Products that can meet both mainstream and counter culture requirements in a quiet and not 'in your face' way, have immense potential of success. In addition, if design solutions could make usage of the cell phone easier in both modes (such as in the 3 counter culture scenarios described), the penetration and adoption rate would be even faster and higher.

4.2 Television

It is interesting to note that one of the most successful media is television and this is for various cultural profiles. What is even more interesting is the kind of 'culture strain' the TV helps deal with. Television continues to be one of the most popular home entertainment media in China, USA and India. Interestingly, in each of these countries, television is used in both 'mainstream' culture as well as 'counter culture' modes. The fact that television reflects attributes of the mainstream culture is very well known. However, the average American and Chinese families use the TV a lot, perhaps because it helps them feel more 'collectivist' when they are in the middle of an individualist environment whether at home or in society (the extreme popularity of chat shows which is essentially about 'talking' to or being connected with people). The average Indian family, on the other hand, use the TV because it provides them an escape route (via suitable programming) from the complete control of the collective (all the films and soap which glorify rebelling against the established societal order).

5 Conclusion - Compensatory Model the Way to Super Hit Opportunity Spaces?

It is interesting to note that the way the three scenarios illustrate the use of the cell phone involve counter culture thoughts or behaviour. amar uses the phone to communicate with his girl friend, deepa uses it to track personal purchases that she hides from her husband and raju gets important information via the cell phone and grins at the many possibilities…

In other words, the cell phone is used by amar, deepa and raju in a compensatory mode.

A compensatory model looks at a given culture and its characteristics. It posits that these characteristics can cause people to behave in a certain predictable way but at the same time the characteristics can put pressure on people. This causes people to behave in accordance with their culture but also seek release to the consequences of the constraints of the culture, at least in subtle ways. This behavior would be classified counter culture.

The questions that need more research, for designers and developers of new products/concepts or for those entering new markets, are:

- Is it possible that those opportunity spaces/concepts that are used in a compensatory mode BUT in a form that is very much in keeping 'with culture' norms, are the potential candidates for major success?
- Should designing to accommodate compensatory mode usage become a standard part of the design process?

These questions can be answered with further research. If the 'compensatory model of product usage' is indeed correct, then it implies a shift in focus from the 'given' cultural characteristics to the 'tensions' between the 'given' and the 'desired'.

Specifically in the emerging economies, where the 'old' and the 'new' exist in startling juxtaposition, for products to be successful, the amars, deepas and rajus must

be empowered to deal smoothly with the duality of their existence. A duality that is captured very well by the Scottish journalist, James Cameron [3]:

> I like the evening in India, the one magic moment when the sun balances on the rim of the world, and the hush descends, and ten thousand civil servants drift home on a river of bicycles, brooding on Lord Krishna and the cost of living.

References

1. Bell, Genevieve: Insights into Asia: 19 Cities, 7 Countries, 2 Years-What People Really Want from Technology. Technology@Intel Magazine (2004)
2. Hofstede, G.: Cultures and Organizaion. McGraw-Hill, New York (1991)
3. Varma, P.: Being Indian. Penguin Books India, New Delhi (2004)

Smart Strategies for Creating Culture Friendly Products and Interfaces

Apala Lahiri Chavan

Vice President - Asia
Human Factors International
Chemtex House, 4th Floor
Hiranandani Gardens, Powai
Mumbai 400 072
apala@humanfactors.com

Abstract. We increasingly live a 'local' global existence, whereby we are affected by the connectedness of the world but at the same time desire to retain our local identity. In this scenario, what strategy should one adopt when designing products and interfaces for use across the world? While we know the pitfalls of the 'one size fits all' strategy, is there an alternative way to include the cultural element in design without incurring huge cost and effort? This paper discusses one such strategy that allows cultural customisation without the' kill bill' budget.

1 To Do or Not to Do…That Is Still the Question!

In this era of the supposed 'global village', there is still much debate about the imperative or otherwise, to create products and interfaces that are a close cultural 'fit' to its users. There are many voices that cite the 'global village' model as a reason to have a one size fits all approach.

However, there is overwhelming research support for the need for cultural customization, emerging from real life experiences of a wide variety of corporations who have tried the 'standardised' strategy approach and have run into major problems across the world.

Professionals from disciplines ranging from anthropology, visual design, usability, product design etc. are now increasingly raising their voice in favour of a rational strategy that would allow products and interfaces to be designed for the cultures where they are going to be used.

Why then is the debate continuing? One of the reasons is the challenge faced by multinational corporations who sell their products across the world and therefore are faced with the ruthlessly practical aspect of the cost and effort involved in cultural customization. Imagine the mega budget involved if Dell or HP or Intel had to totally localize their products and/or their websites for the 90 or more countries that they now reach out to. Creating and maintaining 90+ variations of their products and local sites that communicate information about these local products, would involve a huge budget and managing a major effort on an ongoing basis.

There is no doubt that the most effective kind of cultural customization is when one systematically understands the 'cultural needs' of users in each target culture and

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 27–32, 2007.
© Springer-Verlag Berlin Heidelberg 2007

designs a product/interface that best meets those needs in a culturally familiar way. However, given the high cost associated with this kind of ideal route, and the discouraging effect that has on any initiative to customize culturally, it is worthwhile to explore whether there are other smart strategies to achieve a large part of the cultural customization goal without a large part of the cost.

2 And so to Cultural Dimensions

Cultures can be described according to specific *characteristics* or categorized into *value categories* or *dimensions* of national culture. Dimensions are generally developed from large numbers of variables by statistical data reduction methods (e.g. factor analysis) and provide scales on which countries are scored. Dimensions that order cultures meaningfully must be empirically verifiable and more or less independent'. (De Mooij) There are several categorizations that help in evaluating similarities or differences between cultures. Some of the most commonly used ones are:

Hofstede
Power Distance,
Individualism/Collectivism
Masculinity/Feminity
Uncertainty Avoidance
Long term /Short term orientation

Trompenaars
Universalism versus particularism
Neutral versus affective
Diffuse versus specific
Achievement versus ascription

Hall
High context versus low context
Monochronic versus polychronic
Linear time versus cyclical time

It is a myth that there are universal values that can be used when designing products for the world. Several studies have reiterated the fact that cultures differ and the dimensions mentioned above are one good way of understanding how cultures differ and how users look at the same product in different ways, colored by their 'cultural glasses'.

3 Relating Cultural Dimensions to Users and What They Use – Two Examples

3.1 The Differential Effect of Brand on Asian (Collectivist) Versus American (Individualist) Consumers (Nancy Wong, Bernd Schmitt)

Individualists are likely to value brand image more because a brand's idiosyncratic meanings allow them to create individual and unique relationships with the same

brand and yet maintain a different self –identity from others ("x is my favourite brand because this is what I was using on the day of my promotion').

They will be more disposed to judge each product as an individual.

On the other hand, collectivists are more likely to value things that enhance their relationships with others within the social in groups but elevate their social status to members from the out groups. As a result, they may value brand awareness more for a brand's signifier role in establishing group identity and social hierarchy("y is my favourite brand because it is used exclusively by the elite group").

Collectivists will place more emphasis on the product's affiliation to a group, such as a brand, manufacturer, or country of origin.

3.2 Culture Explains (de Mooij1997, 1998, 2000, 2001)

While for some products differences between countries worldwide can be explained by differences in national income, in more economically homogeneous Europe most differences can only be explained by culture.

Differences in media usage are persistent because the media are part of countries' culture. Although, for some media, differences worldwide are related to national income, in the developed world and Europe in particular, differences in radio ownership can only be explained by culture (Fig. 1.).

The number of radios per 1,000 population is correlated with individualism. This correlation becomes more significant over time. In individualist cultures everyone has his/her own radio, while in collectivist cultures one radio per family is enough.

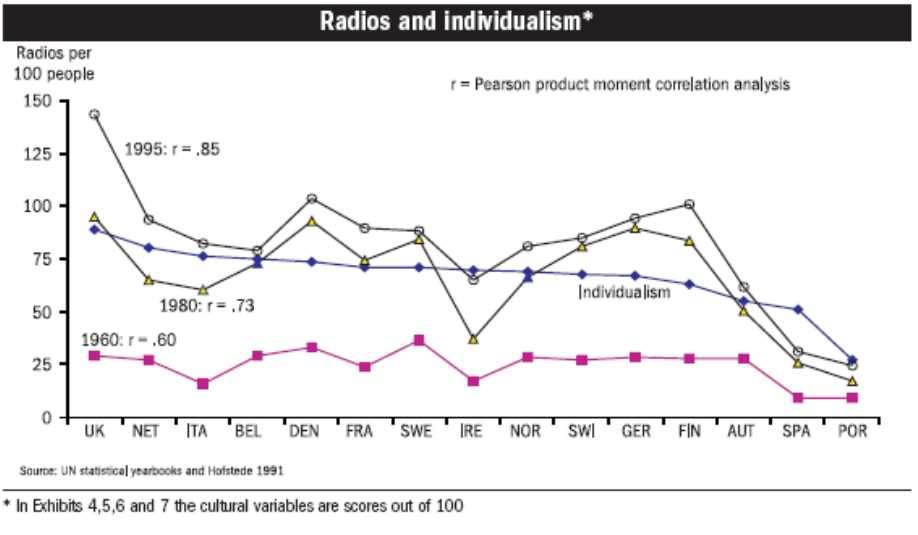


Fig. 1. Relationship between the dimension of 'individualism' and ownership of radio sets in Europe

Many other consumption differences can be predicted and explained by analyzing the relationship between consumption and scores on Hofstede's dimensions of national culture. For example, culture has been shown to influence the volume of mineral water and soft drinks consumed, preferences for new or used cars, ownership of insurance products, possession of private gardens, readership of newspapers and books, television viewing, ownership of consumer electronics, use of the Internet, use of cosmetics, deodorants, toothpaste and hair care products, and consumption of fresh fruit, ice cream and frozen food as well as numerous other products and services.

4 And so, Can This Lead to a Smart Strategy?

If cultural dimensions can predict consumption and usage behavior, then would it not be useful to FIRST group the countries one aims to reach out to, in clusters of countries with similar scores for each dimension? Scores for Hofstede's five dimensions of national culture are available for 59 countries (Fig 2.). The dimensions are measured on index scales from 0 to 100. The dimensions are Power Distance (PDI), Individualism/Collectivism(IDV), Masculinity/Femininity (MAS), Uncertainty Avoidance(UAI), and Long-Term versus Short-Term Orientation(LTO).

For example, if we take a look at the index of Hofstede's dimensions and select countries whose rank on the individualism dimension is low (rank of 30 or lower), we get an interesting mix of countries (Fig. 2.). The score on the individualism dimension being low implies that the countries are collectivistic (the opposite of individualistic). How does hofstede define collectivism?

Collectivism (Demooij)
In *individualistic* cultures, people look after themselves and their immediate family only and want to differentiate themselves from others. There is a need for privacy. In *collectivistic* cultures people belong to in-groups who look after them in exchange for loyalty. People prefer to conform to the norms adopted by others instead of differentiating themselves from others.

In individualistic cultures the person is viewed as an independent, autonomous entity with a distinctive set of attributes, (traits, abilities, motives and values). In collectivistic cultures individuals are fundamentally dependent on each other. The self cannot be separated from others and the surrounding social context. Self-reflection is more common among individualists than collectivists because for the latter their relationships to others are more important than self-knowledge.

As it appears from the index, Asians, Latin Americans and Africans are collectivists as opposed to North Americans who are individualists.

In other words, all these countries that scored high as collectivist cultures have that dimension as a significant common aspect of their culture, in spite of the geographical distance from each other.

Taking advantage of this similarity, a company like HP could decide to design one set of products and one website for this group of 22 countries, instead of 22 different products and websites for each of these 22 countries.

PDI Power distance index
IDV Individualism index
MAS Masculinity index
UAI Uncertainty Avoidance index
LTO Long-Term Orientation index

	PDI		IDV		MAS		UAI		LTO	
	rank	score	rank	score	rank	score	rank	score	rank	score
Arab Countries	7	80	26/27	38	23	53	27	68		
Argentina	35/36	49	22/23	46	20/21	56	10/15	86		
Australia	41	36	2	90	16	61	37	51	15	31
Austria	53	11	18	55	2	79	24/25	70		
Bangladesh									11	40
Belgium	20	65	8	75	22	54	5/6	94		
Brazil	14	69	26/27	38	27	49	21/22	76	6	65
Canada	39	39	4/5	80	24	52	41/42	48	20	23
Chile	24/25	63	38	23	46	28	10/15	86		
China									1	118
Columbia	17	67	49	13	11/12	64	20	80		
Costa Rica	42/44	35	46	15	48/49	21	10/15	86		
Denmark	51	18	9	74	50	16	51	23		
East Africa	21/23	64	33/35	27	39	41	36	52		
Ecuador	8/9	78	32	8	13/14	63	28	67		
Finland	46	33	17	63	47	26	31/32	59		
France	15/16	68	10/11	71	35/36	43	10/15	86		
Germany FR	42/44	35	15	67	9/10	66	29	65	14	31
Great Britain	42/44	35	3	89	9/10	66	47/48	35	19	25
Greece	27/28	60	30	35	18/19	57	1	112		
Guatemala	2/3	95	53	6	43	37	3	101		
Hong Kong	15/16	68	37	25	18/19	57	49/50	29	2	96
India	10/11	77	21	48	20/21	56	45	40	7	61
Indonesia	8/9	78	47/48	14	30/31	46	41/42	48		
Iran	29/30	58	24	41	35/36	43	31/32	59		
Ireland (Republic of)	49	28	12	70	7/8	68	47/48	35		
Israel	52	13	19	54	29	47	19	81		
Italy	34	50	7	76	4/5	70	23	75		
Jamaica	37	45	25	39	7/8	68	52	13		
Japan	33	54	22/23	46	1	95	7	92	4	80
Malaysia	1	104	36	26	25/26	50	46	36		
Mexico	5/6	81	32	30	6	69	18	82		
Netherlands	40	38	4/5	80	51	14	35	53	10	44
New Zealand	50	22	6	79	17	58	39/40	49	16	30
Nigeria									22	16
Norway	47/48	31	13	69	52	8	38	50		
Pakistan	32	55	47/48	14	25/26	50	24/25	70	23	0
Panama	2/3	95	51	11	34	44	10/15	86		
Peru	21/23	64	45	16	37/38	42	9	87		
Philippines	4	94	31	32	11/12	64	44	44	21	19
Poland									13	32
Portugal	24/25	63	33/35	27	45	31	2	104		
Salvador	18/19	66	42	19	40	40	5/6	94		
Singapore	13	74	39/41	20	28	48	53	8	9	48
South Africa	35/36	49	16	65	13/14	63	39/40	49		
South Korea	27/28	60	43	18	41	39	16/17	85	5	75
Spain	31	57	20	51	37/38	42	10/15	86		
Sweden	47/48	31	10/11	71	53	5	49/50	29	12	33
Switzerland	45	34	14	68	4/5	70	33	58		
Taiwan	29/30	58	44	17	32/33	45	26	69	3	87
Thailand	21/23	64	39/41	20	44	34	30	64	8	56
Turkey	18/19	66	28	37	32/3	45	16/17	85		
Uruguay	26	61	29	36	42	38	4	100		
USA	38	40	1	91	15	62	43	46	17	29
Venezuela	5/6	81	50	12	3	73	21/22	76		
West Africa	10/11	77	39/41	20	30/31	46	34	54		
Yugoslavia	12	76	33/35	27	48/49	21	8	88		
Zimbabwe									19	25

Fig. 2. Index showing the rank and score of countries, for the five Hofsetdian dimensions

This would allow HP to explore all aspects of collectivism and how products designed for collectivists need to be different and how the website needs to communicate differently, from one designed for individualists.

It would, for example, be obvious that the product/s would need to allow shared usage and make it possible for the user to feel a sense of belonging/affiliation to the 'collective' or 'in group'. Further, the product would NOT make the user stand out or flaunt their individuality. It would, in fact, allow the opposite to happen.

The website for these countries would emphasise the shared usage capability and communicate 'in group' acceptance by ownership of the product.

The bottom line for HP would be a major win- win, in having created a few products and one website (thereby saving cost and effort) BUT all of these would be much more culturally customized (and hence more attractive to the users) than if they had used a one size fits all strategy and achieved cost saving.

5 Conclusion

As we realize that the 'world is flat', it becomes evident that in this flat world , cultures are increasingly visible to each other. However, that does not imply that cultures are converging into a truly homogenous global village. On the contrary the sharp juxtaposition of different cultures against each other is making people live life as an interesting double act. When in a 'global' environment (such as traveling outside ones country), there is evidence of 'global homogenous' behavior, but on return of the native there is evidence of a desire to recharge oneself with local 'flavors'. In this midst of this alternating reality, organizations wanting to reach out to the world, with their products and services, have no choice but to explore smart strategies that allow them to step closer to their users but also remain competitive as a business.

References

1. Hofstede, G.: Cultures and Organizaion. McGraw Hill, NY (1991)
2. Edward, H.: Beyond Culture. Doubleday. Anchor Books
3. de Marieke, M., Geert, H.: Convergence and divergence in consumer behavior: implications for international retailing. Journal of Retailing 78, 61–69 (2002)
4. de Marieke, M.: Convergence and divergence in consumer behaviour. World Advertising Research Center (2001)
5. Fons, T., Alfons, T., Charles, H.-T.: Riding The Waves of Culture: Understanding Diversity in Global Business. Mc-Graw Hill

When in Rome… Be Yourself: A Perspective on Dealing with Cultural Dissimilarities in Ethnography

Apala Lahiri Chavan[1] and Rahul Ajmera[2]

[1] Vice President - Asia
Human Factors International
Chemtex House, 4th Floor
Hiranandani Gardens, Powai
Mumbai 400 072
apala@humanfactors.com
[2] Project Manager
Human Factors International India Pvt. Ltd.
310/6, H.R. Complex, 2nd Floor,
Koramangala, 5th Block,
Bangalore - 500 095
Tel.: +91 80 4150 7221/2/3
Fax: +91 80 4150 7220
rahul@humanfactors.com

Abstract. With the 'flattening' of the world, increasingly, our design research teams are called upon to execute projects in cultures that are foreign to them. Design research involves deep dive ethnography that needs to be carried out in a relatively short span of time. It is in these design ethnography studies that we have realized the impact of cultural difference between the researchers and the researched. This paper attempts to discuss our findings on the subject.

1 Introduction

Human Factors International Inc. (HFI) is a 240 people, $20 million consulting practice working in the area of user centred design, with a mission to improve the interactions that people have with computers and other digital systems. HFI offers end to end solutions for Web/Intranet and Internet-based applications, Software Applications, IVR Systems, Handheld Devices, Telemetric, Public Service Networks, Medical and Automation Equipments and help make our clients' existing offerings more user centric, optimized and efficient. In the wake of the recent interest in research and development for business innovations for the emerging markets and products & services for new markets, we have established ethnography and design research as a service area along with the existing areas of HFI's activities & services. HFI's interest in this area is reflected through successful collaborations with research initiatives launched by global corporate as well as academic institutions, e.g. HP Labs, Nokia Research, NCR, Media Lab Asia, Intel etc.

2 Role of Ethnography in Innovation

As technology permeates each and every aspect of our lives, we are constantly faced with situations where there are mismatches between its role and our lives. It is now

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 33–36, 2007.
© Springer-Verlag Berlin Heidelberg 2007

known even to technologists that the power of technology can only be gauged by its social relevance and acceptability. Understanding how users experience technologies necessitates a concern with social and cultural meaning; what does the product mean to the user; what does it mean in the context of particular cultures, what does it mean in terms of its broad social and global environment? [1]. Socio cultural relevance is not limited to technological products or services but it is given a lot of importance as development costs associated with development of technologies are typically very high and the development cycles very long. These high stakes make it imperative that a reliable requirement capture method be implemented.

2.1 Ethnography: Introduction

Ethnography, is a traditional method that belongs to the fields of sociology and cultural anthropology. It involves the study of people performing activities and interacting in complex social settings in order to obtain a qualitative understanding of these interactions [2]. Classic ethnographic analysis is typically based on extensive interviews, observations and field studies that consume months or years. This immersion in the world of study participants permits the investigator to create a detailed, fine grained picture of a culture. The time intensive nature of these conventional ethnographic studies however rarely ever seem to fit corporate cultures and their technology development time lines. In addition, the discursive, qualitative descriptions necessary to present this rich picture does not often align well with formal system specifications [3]. Nonetheless the benefits of examining field situated user activity remains inviting [4].

In response, classic ethnographic studies have been modified to accommodate commercial and educational circumstances, to allow practitioners and students to gain insights more quickly and more directly from potential end users.

Our studies typically include depth interviews, cultural probes, contextual inquiry, observational studies etc. The techniques and methods we employ are modified versions of methods borrowed from traditional ethnography. We at HFI refer to our overall innovation process as “Contextual Innovation”. While we effectively employ ethnography based methods it is also important for us to realize that there are trade offs with this resource effective industry driven version of ethnography. Though the scope of this paper does not permit discussion of the short falls but some of the issues would be raised as they are directly related to the focus of the paper, that is, understanding the effect of cultural differences on ethnography.

3 Cultural Differences

As we move from project to project we have conducted studies in a variety of different socio-cultural contexts varying sometimes from a household in New York to one in rural India. In a world that is rapidly globalizing, the instances of this are only going to increase for user researchers around the world. In many cases our researchers have had to conduct their studies in cultures fairly alien to them. It is during these studies that we have come to realize the importance and the impact of these cultural differences. In this paper we will try to discuss areas where we have realized these differences play a major role and how they can be addressed.

3.1 Addressing Cultural Differences: Traditional Ethnography vs Design Ethnography

One of the important differences between our kind of research vs traditional ethnography is the time the researchers spend on the field. Typically our researchers segment the study into focused visits during which they attempt to unearth the most relevant aspects of the contexts. These visits are not long enough for the researcher and the researched to be completely at ease. 'There is an assumption that as the researcher becomes a more familiar presence, participants are less likely to behave uncharacteristically".

While we would like to believe that our attempts at making the participants comfortable, work, but in doing so the dynamics still cannot always be compared to that in traditional ethnography. This implies that when researchers from foreign cultures enter an unknown context for short periods(as is mostly the case with design ethnography) the effect of this cultural difference may impact the study adversely .

One clear impact on the study in such cases is that researchers themselves stand out in the context. This point is of great interest because there are a lot of researchers who stress on trying to "merge into the researched contexts". Their prime reason of advocating such an approach to handle "the public glare" is to make the participants comfortable. We have in some cases tried to follow this approach but have realized that, sometimes, in spite of our best efforts we could not "merge in" and trying to awkwardly deal with this inability to immerse ourselves in the context of our participants only made us feel more uncomfortable. In our opinion, rather than struggle with an impossible immersion, the better approach is to embrace the 'non-immersion' and to actually use the 'foreignness' of the researcher to our advantage.

After all, in ethnography, the researcher deliberately constitutes himself as the "other" in embarking on the enterprise of fieldwork. Having become the 'other', the researcher, in classical ethnography, tries to make sense of the 'lived' experience of the 'people' he is trying to understand. The 'key informant' from amongst the 'natives' plays a critical role in helping the researcher understand this 'lived experience.

When we say that using the 'foreignness' is often more helpful rather than trying to immerse ourselves (as researcher) into the 'lived' experience, one primary reason is that we have, in this abridged form of ethnography, 6 days available rather than 6 months. Hence, while we very much constitute ourselves as the 'other', what we do differently is that often we do not go in for immersion in the context of our participants.

And this leads us to the much debated concepts of 'emic' and 'etic' knowledge.

4 Neutrality vs Immersion

'A researcher who works with emic knowledge, will look into – and generally accept unproblematically – the rules, terms, reference points and logic of the person she is studying. Part of what she will convey to readers of her research is this internal system of logic of the group or person, and her conclusions will derive from that. The

analogy to a phoneme is clear – in linguistics it denotes a meaningful unit of sound specific to a particular language.

An etic researcher will ask her informant questions based on her own perspective and concerns, which are often seen to be 'scientific', or 'universal'. Phonetics, which discusses sounds *qua* units of sound, rather than sounds in context, is a fitting theoretical analogy. The researcher will present an interpretation of her data that draws conclusions using external categories, valuations, and judgments. In the social sciences today claims of scientific methods, universalism, and neutrality are heavily contested.

A simple way of thinking about the distinction is this: an emic researcher will 'go native' to some extent, behaving, speaking, eating, and thinking like her subjects of study. An etic researcher will stay on the edges, assessing them on her own terms.'[6].

5 Conclusion - Stranger in a Strange Land!

When working with abridged ethnographic methods, as in design research, the 'etic' approach often elicits more open and honest responses from the participants. The feeling that participants get, of, 'oh this chap is a foreigner and therefore its ok that he is asking such strange/stupid questions' makes it much easier for us to ask questions that would normally be thought of as 'awkward' or even a 'strict no-no' and equally easy for the participants to answer what would otherwise be considered embarrassing or very personal questions.

Moreover, the process we follow when working on contextual innovation projects places considerable emphasis on understanding the clients ecosystem. Thus, while we might be foreign to the context, this emphasis gives us a framework to probe from the 'outside' if necessary and be familiar with the participants' "perspective". It is this understanding of the perspective that is taken into account in our etic "Point of view".

As Pawan Verma says [7], 'societies reveal how they actually think and behave in the smallest things. Behavioural patterns have to be discovered not in the considered stance before an observer, but in the insignificant reflex preceding or following it'.

Being an 'outsider' trained in design ethnography, it often becomes easier to assume that there IS a 'considered stance' and therefore be on the lookout for the 'insignificant reflex'.

References

1. Bell, G., Blythe, M., Gaver, B., Sengers, P., Wright, P.: Designing Culturally situated Technologies for the home. CHI 2003 (2003)
2. McCleverty, A.: Ethnography. Computer science 681: research methodologies (1997)
3. Hughes, J., King, V., Rodden, T., Anderson, H.: The role of ethnography in interactive systems design. Interactions (1995)
4. Millen, D.: Rapid ethnography: Time deepening strategies for HCI
5. Anderson, R.: Representations and requirements: The value of ethnography in System design. HCI 1992 (1992)
6. http://www.articleworld.org/index.php/Emic_and_etic
7. Varma, P.: Being Indian. Penguin Books India, New Delhi (2004)

Designing User Interfaces for Mobile Entertaining Devices with Cross-Cultural Considerations

Chien-Hsiung Chen[1] and Chia-Ying Tsai[2]

Graduate School of Design, National Taiwan University of Science and Technology
43 Keelung Road, Section 4. Taipei, 106, Taiwan
[1] cchen@mail.ntust.edu.tw, [2] virginiatsai@gmail.com

Abstract. The purpose of this study is to explore the design process regarding how interaction designers in Taiwan deal with the OEM and ODM types of product and user interface design styles pertinent to mobile entertaining devices, such as MP3 players and portable media players (PMP). In addition to the discussion of what culture is and the way to design international user interfaces with cross-cultural considerations, detailed interaction design process with real world design examples is also introduced. It is hoped that the design process mentioned in this paper can be a good reference to interaction designers when they design product and user interface to satisfy users of various cultural backgrounds.

Keywords: Mobile entertaining device, Cross-cultural design, Interaction design, Usability testing.

1 Introduction

"The world is flat." As Friedman [2] points out that the physical boundaries among the world economic entities are disappearing. Designing products for international users all over the world will be the goal for future marketing strategies. Traditionally, the product industry in Taiwan has long been operated as the Original Equipment Manufacturing (OEM) and Own Designing & Manufacturing (ODM) types of design styles. Only few corporate companies were able to create their product brand names and conduct their product design process as the Own Branding & Manufacturing (OBM) style. In addition, due to the progress of advanced digital technology, the product lifecycle for mobile entertaining devices, such as MP3 players and portable media players (PMP), has been decreased because new products are introduced to the market daily. Many of these mobile entertaining devices were designed and manufactured in Taiwan based on the OEM and ODM types of product and user interface design styles for users of various cultural backgrounds. Therefore, the time required for product and user interface design and manufacture before it is released to the world market has been significantly reduced because of global competitions. The earlier a company introduces its innovative product to the world market, the more likely that this company may have a better chance to occupy a bigger portion of the product's market shares and may be able to lead the future development of this

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 37–46, 2007.
© Springer-Verlag Berlin Heidelberg 2007

product. Therefore, it is important for an interaction designer to create an internationalized product and user interface with cross-cultural considerations to accommodate the majority of the international users' requirements.

The purpose of this study is to explore the design process regarding how interaction designers in Taiwan deal with the OEM and ODM types of product and user interface design styles pertinent to the mobile entertaining devices, such as the MP3 players and portable media players (PMP). Based on the summaries from interaction designers working for the major corporate companies designing and producing mobile entertaining devices in Taiwan, the design process pertinent to the OEM and ODM types of product and user interface design styles are discussed in this paper.

2 What Is Culture?

Culture can be viewed as "shared patterns of behavior" [5]. A cultural environment is be able to provide an individual with an emotional space in which set of beliefs, values, and behaviors can be commonly shared by all the members within the same society or ethnic group [1]. Cultural traditions (i.e., patterns) must be generally agreed upon by the majority of the members of the culture, not just by an individual alone. Therefore, within one culture, the majority of the members will share the same image perceptions pertaining to the value or even the interaction style of the mobile entertaining devices. However, if the same mobile entertaining devices are designed to be used among different cultures, more cross-cultural design considerations will need to be conducted to guarantee the product's success.

Vaske and Grantham [8] point out three basic characteristics of culture: (1) Culture is generally adaptive. It is generally adaptive to the particular conditions of both physical and social environments. (2) Culture is mostly integrated. It is mostly integrated in that the elements or features which make up the culture are mostly adjusted to or consistent with each other. (3) Culture is always changing. It is always changing because of adapting to certain cultural events or integrating with other cultures. An interaction designer should fully understand the characteristics of culture before s/he can design a product and user interface for users form different cultures.

Culture can also be viewed as communication [3]. That is, within one culture, all the members are able to interact with each other based on similar cultural behaviors. Hall [3] organized cultures by amount of information implied by the setting or context of the communication itself, regardless of the specific words spoken. He argued that cultures differ on a continuum ranging from high to low context. In high-context cultures, the communication is implied by a physical setting or by an individual's beliefs and values. Information is shared among all members of the same culture, but some have more privileged access than others. For example, Japanese, Mexican, and African-American cultures are all related to high-context cultures. In low-context cultures, the communication among culture members is expected to be explicit, and everyone has equal access to available information. Examples of low-context cultures include German, Swedish, and European-American cultures. In the context of cross-cultural design, the communication between a human and a product or a user interface is moving from low-context to high-context interaction. This can be due to that fact that the progress of digital computing technology has made the traditional rigid

control on a product or a user interface no longer necessary. Instead, intelligent microchips in the 21st century have enabled multiple and flexible interaction styles to facilitate the user's interactions. That is, the intelligent product is able to sense the user's task intentions and automatically execute the functions for the user.

3 Designing International User Interfaces with Cross-Cultural Considerations

To an interaction designer, the purpose of conducting international user interface design is to create useful and effective user interfaces which can be utilized by all the potential users with various cultural backgrounds. In fact, international user interface design should be considered as a cross-cultural collaborative work between interaction designers and users from different cultures [4]. Designing international user interfaces requires taking the concept of both internationalization and localization of user interfaces into account. Internationalization is the process of designing a base user interface which can be further integrated with various cultural factors to meet with different cultural needs. Localization is the process of adapting an internationalized user interface based on the features of a particular culture. In fact, the process of interface internationalization will facilitate the process of interface localization as well. The process of interface internationalization can provide a dynamic framework (i.e., the structure) in which interface localization can be implemented by adding cultural factors into the design. Because the internationalization of user interfaces requires intensive cross-cultural design considerations, an interaction designer will need to identify and separate basic principles regarding user interface design into culturally independent and culturally dependent variables. The culturally independent variables are the variables used to help interface internationalization, and the culturally dependent variables can be used to facilitate interface localization.

4 The OED and ODM Design Process in Taiwan

Because of the OEM and ODM types of product and user interface design styles in Taiwan, most of the product and user interface design projects require interaction designers to complete the design process within three to six months. Otherwise, the proposed products may not be able to occupy a vital place in the international market. Therefore, interaction designers working on this type of design project will need to construct a unique design process to ensure the deadline can be met. This unique design process may include seven stages described as follows:

4.1 Understand the Design Goal

Because of the time constraint, the interaction designer will need to fully understand the goal of the product or the user interface design in order to generate suitable design concepts. To do this, the interaction designer needs to understand three design issues, i.e., the user of the product, the function of the product, and the use environment of the product.

The User of the Product. When conducting the OEM and ODM design project, an interaction designer first needs to know who the target users are. This is because different user groups may have different physical and psychological requirements towards that product or user interface. Their perceptions on the graphical user interface (e.g., icons and menu designs) and interaction styles on the solid user interface (e.g., buttons and switches) may be different as well. In addition, target users' general characteristics are also important for the design considerations. For example, the color used on the display heading of a mobile entertaining device may adopt the "matured colors" style if it is designed for the middle-aged business users (see Fig. 1). On the contrary, if the mobile entertaining device is designed specifically for teenagers, "vivid colors" tend to be used on the display (see Fig. 2).



Fig. 1. Matured colors are adopted for middle-aged business users



Fig. 2. Vivid colors are used for teenagers

The Functions of the Product. The functions provided in the mobile entertaining devices will affect how users interact with the user interface. For instance, the function of global positioning system (GPS) is often incorporated within this type of mobile entertaining device. In addition to provide users with precise position information, other design factors, such as battery capacity, fall protection, and water protection, are also important and an interaction designer needs to take these design factors into serious account.

The Use Environment of the Product. Depending on the user's work environment, an interaction designer will need to consider if this mobile entertaining device will be designed by adopting touch pen interaction style or just allow users to interact with the interface by using their finger tips. If it is designed for touch pen interaction, more function icons and detailed operation icons can be provided on the display because users can use a touch pen to conduct a more precise interaction (see Fig. 3). On the other hand, if this device allows user to interact with its user interface by using their finger tips, less function icons should be provided on the display and the size of the function icons should be larger for easier interactions (see Fig. 4).

4.2 Plan for Systematic and Series Designs

Because of the OEM and ODM characteristics, the product development cycle and lifecycle are very limited in Taiwan. To an interaction designer, the design goal is to put the product into the market as early as possible without sacrificing its quality. Once the product is in the market for around three to six months, a new generation of that similar product will replace its market position. Because this new generation is not very much different from its predecessor and only slightly changes have been

made, an interaction designer will need to plan this type of design strategies in advance in order to create new products based on systematic and series considerations. By so doing, the interaction designer not only can minimize the production cost, but also can control the time frame for designing a new product. For example, Fig. 5 is the display showing the original functions of a mobile entertaining device. Fig. 6 is the new design illustrating more functions than its predecessor. The interaction designer should be able to complete designing the new generation within the possible shortest time.

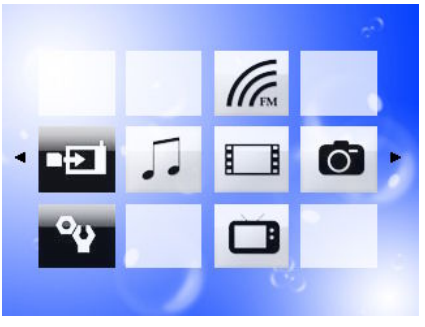


Fig. 3. The display with smaller icons designed for touch pen interaction



Fig. 4. The display with larger icons designed for finger tips interaction



Fig. 5. The original design showing the functions of a mobile entertaining device

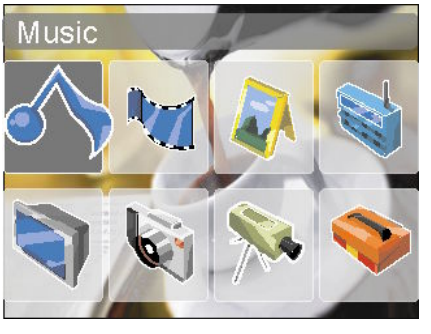


Fig. 6. The new design illustrating more functions than its predecessor

4.3 Communicate Well with Other Design Teams

The product and user interface design process in the current OEM and ODM design industry tends to be a team work. That is, an interaction designer cannot complete the design by himself/herself. That is, s/he needs to work with project manager, electrical engineer, mechanical engineer, software engineer, and other stakeholders. In order to complete the design within the scheduled time, all the parties need to communicate well. Knowing other team members' requirements in advance will also ensure the quality of the design. For example, when an electrical engineer is testing the display quality of a mobile entertaining device, the interaction designer may provide him/her

with various styles of icon designs to be shown on the display (e.g., straight and curve borders, black and white, gray, and full colors, and degree of complexity). Fig. 7 demonstrates the result of an icon shown on an 8-bit display which can present 256 colors. The gradation quality is much better than that on a 4-bit display illustrated in Fig. 8. This is because different display quality may be equipped with different resolutions and limitations as related to price differences. What the engineer wants to achieve is to find out the combination of best presentation quality and lowest display cost to help win the product's price competition.



Fig. 7. An icon shown on an 8-bit display presenting 256 colors



Fig. 8. An icon shown on a 4-bit display presenting only 16 colors

4.4 Conduct User Interface Design

Once an interaction designer understands the client's design specifications and the user's requirements, s/he will start conducting the user interface design. It is an iterative design process emphasizing on the design and testing of generated ideas. During this stage, three types of design variations should be kept in the interaction designer's mind, i.e., hardware variation, content variation, and structure variation, to help achieve best design quality. Detailed explanations are provided below.

Hardware Variation. The interaction designer should always keep in mind that the product and user interface that s/he is currently working on may just exist in the market for a short period of time. Therefore, it is very important to prepare the new generation of the product and user interface in advance. Sometimes the new generation may be different from its predecessor in hardware requirements. For example, the original mobile entertaining device may be designed to be used on a 3.5" display (see Fig. 9). Nonetheless, the new generation may be used on a 7" display. The interaction designer should maintain the user interface design flexibility so the original design can be easily modified within the possible shortest time with the same display quality (see Fig. 10).

Content Variation. The design strategy for the OEM and ODM types of product and user interface design styles is to constantly provide new products on the market to attract users' attentions and, at the same time, encourage them to purchase these new products. Therefore, an interaction designer needs to be aware that the new generation may just have minor changes to its predecessor in order to save the cost of

developing a whole new product or user interface. Most of the time, the minor change may just be the addition of a new function to the existing product and/or fixing the current problems. If the software of the user interface is designed with object-oriented considerations, it will not be too difficult for a software engineer to modify the coding of the existing software.

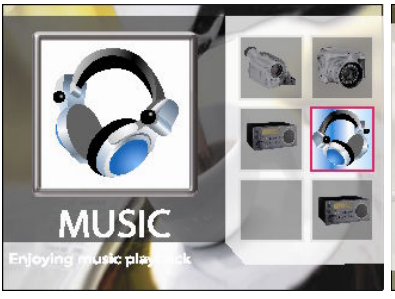


Fig. 9. The mobile entertaining device designed with a 3.5" display

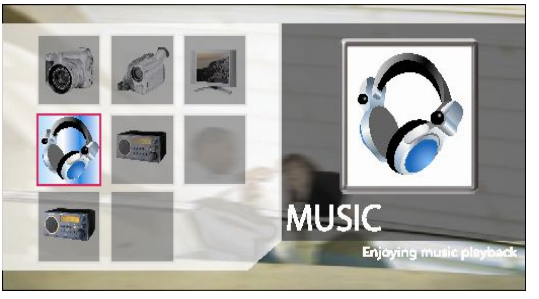


Fig. 10. The mobile entertaining device designed with a 7" display

Structure Variation. The process of modifying the structure of a user interface can be very complicated and sometimes very difficult. Very often, the interaction designer may need to start the design process all over again. For example, Fig. 11 shows that the function icons on the main menu of a mobile entertaining device can be rotated and controlled by two arrow buttons on the sides. Fig. 12 illustrates that the function icons on the main menu can be chosen by touch-sensitive control style. Though these two user interfaces look similar, they are created based on two different interaction styles. It will be very difficult to convert one design based on the other. Therefore, it is important to obtain users' viewpoints in advance and inform the client which design can best satisfy most uses' interaction styles as soon as possible. Once the design decision has been made, try not to modify the design again for it can be very time and resources consuming.

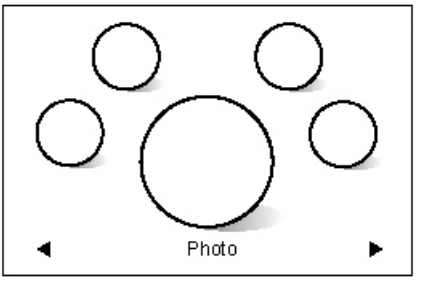


Fig. 11. The functions on the main menu can be rotated and controlled by two arrow butons

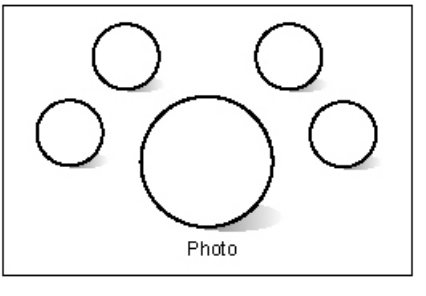


Fig. 12. The functions on the main menu can be chosen by touch-sensitive control style

4.5 Implement Cross-Cultural Design Considerations

There are two design strategies to be used before the product is put into the market, i.e., design for the general public and design for a specific target user group. These two user groups may have their own cultural characteristics. Very often, the strategy of designing for the general public will be adopted when the product is first introduced to the market. The purpose is to draw the public's attention and by so doing can also promote the product's brand name image. After that, limited editions with minor modifications (i.e., textures, colors, or endorsements from a famous person) of the same product will be introduced to the market to prolong the product's lifecycle. In order to conduct the second strategy, an interaction designer should be very aware of the target users' cultural features so that the limited edition can attract their attentions. Furthermore, in order to achieve the goal of conducting product variation, the concept of module design will be considered beforehand. That is, the changeable product or user interface elements should be designed in the forms of flexible modules. Therefore, the new generation of the product can be modified with less efforts and costs but still can be designed with a fresh new look. For instance, Fig. 13 shows the icon designs using black and white colors and simply style that can be used on a less expensive mono-colored display. Fig. 14 illustrates the icons designed by adopting more complex lifestyle images. This type of design is often used in the Asian market because of cultural characteristics. Fig. 15 demonstrates the icons designed by adopting the image of glassware in Chinese culture to help promote the quality of the product and user interface.



Fig. 13. The icons designed with black and white colors and simple style to be used on a mono-colored display



Fig. 14. The icons designed by adopting more complex lifestyle images and are often used in the Asian market



Fig. 15. The icons designed by adopting the image of glassware in Chinese culture to help promote the quality of the product and user interface

4.6 Construct User Interface Prototypes

Because of the OEM and ODM characteristics, up to now, the interaction designer has spent a lot of time in designing the product and user interface. The time tends to be running out and s/he may not have enough time to construct the user interface prototype for testing purpose. The interaction designer may just spend one day to ask his/her colleagues or someone else working in a nearby office to act as a user to help provide opinions. After a brief modification, the interaction designer may transfer the user interface design to the software engineer for coding process. It is very likely that the user interface design may still contain potential interaction problems. The interaction designer and software engineer will need to jointly solve these unfound problems along the coding process. Nonetheless, if there is time for constructing user interface prototypes, two types of prototypes can be made during the design process, i.e., low-fidelity prototype and high-fidelity prototype. Fig. 16 shows the low-fidelity prototype to be used for design discussions. Fig. 17 illustrates high-fidelity computer simulation prototype to be used for usability testing to help acquire information regarding user preference and performance.

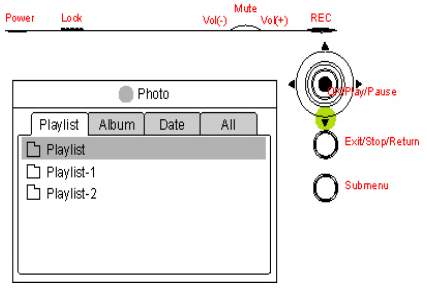


Fig. 16. Low-fidelity prototype used for design discussions

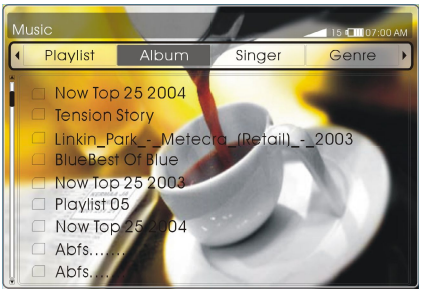


Fig. 17. High-fidelity computer simulation prototype used for usability testing

4.7 Conduct Interface Usability Testing

The International Standards Organization (ISO) defines usability as the effectiveness, efficiency, and satisfaction with which specified users can achieve specified goals in particular environments (ISO DIS 92411-11). Usability testing can be conducted by means of an interface prototype to assess the usefulness of an actual design. The overall goal of usability testing is to identify usability deficiencies existing in the proposed design before its release. The intention is to ensure that the new design will be very easy to learn and use, and that it can provide various functions valued by a target user group or users with various cultural backgrounds.

Under an ideal situation, the process of testing the product and user interface usability should be conducted in the target users' cultural environment in order to obtain the first hand information. However, in Taiwan, because of the time and resources constraint, the interaction designer may not be able to conduct a full scale usability testing by recruiting real users from oversea. Not to mention that s/he may need lots of time to conduct the experiment and perform data analysis. Therefore, most of the

time, the technique of heuristic evaluation will be conducted at the interaction designer's company. That is, the interaction designer may invite 3 to 5 expert users with different design backgrounds (e.g., product design, Website design, graphic design, etc.) to take part in the usability testing process. These experts may spend 1 to 3 days playing with the product and the user interface. According to Nielsen [6][7], most of the major design problems can be identified by these experts. After that, the interaction designer may still have some time to co-work with the software engineer for the last stage modifications before the product is released to the market.

5 Conclusion

Because of the OEM and ODM types of product and user interface design styles in Taiwan, the time required for the design development is strongly constrained. In order to compete and survive in the international market, an interaction designer will need to construct his/her own unique design process to fit in this rapidly changing environment. This research study demonstrates the unique product and user interface design process based on cross-cultural considerations with real world design examples. It is hoped that this unique design process can be a good reference for interaction designers to help design product and user interface that can satisfy users of various cultural backgrounds.

Acknowledgments. Financial support of the research by National Science Council under the grant NSC 95-2221-E-011-046 is gratefully acknowledged.

References

1. Ember, C.R., Ember, M.: Anthropology. Prentice-Hall, Englewood Cliffs, NJ (1977)
2. Friedman, T.L.: The World is Flat: A Brief History of the Twenty-First Century. Farrar, Straus and Giroux, New York (2005)
3. Hall, E.T.: The Hidden Dimension. Doubleday, New York (1969)
4. Ito, M., Nakakoji, K.: Impact of Culture on User Interface Design. In: del Galdo, E.M., Nielsen, J. (eds.) International User Interfaces, John Wiley & Sons, New York (1996)
5. Mead, M.: Coming of Age in Samoa. Modern Library, New York (1953)
6. Nielsen, J.: Usability Engineering at a Discount. In: Salvendy, G., Smith, M.J. (eds.) Designing and Using Human-Computer Interface and Knowledge Based Systems, Elsevier, Amsterdam (1989)
7. Nielsen, J.: Big Paybacks from 'Discount' Usability Engineering. IEEE Software 7(3), 107–108 (1990)
8. Vaske, J.J., Grantham, C.E.: Socializing the Human-Computer Environment. Ablex, Norwood, NJ (1990)

Kansei Design with Cross Cultural Perspectives

Kuohsiang Chen, Shu-chuan Chiu, and Fang-chyuan Lin

Department of Industrial Design, National Cheng Kung University, 1 University Road, Tainan 701, Taiwan
kchen@mail.ncku.edu.tw

Abstract. This study aimed to explore the cross cultural perspectives (including that of Taiwan, China, Japan and Korea) toward Kansei design using mobile phone as an example. Formal features, Kansei adjectives and the relationships between them were investigated via Kansei engineering procedures: (1) collecting mobile phone samples and Kansei words; (2) selecting mobile phone samples and Kansei words using KJ method and Factor Analysis respectively; (3) designing four sets of bilingual questionnaires with 5-point Licker Scale; (4) conducting experiments on four sites with questionnaire; (5) analyzing results using Quantification Type I. The achieved tasks include: (1) The Kansei needs of consumers from different culture background; (2) The preferred formal features of a mobile phone among different cultural background; and (3) The relationships between Kansei words and formal features for different cultural background. The results can be used as reference for designing cross-culture mobile phones as well as other closely related products.

Keywords: Cross-cultural, Culture difference, Formal features, Kansei engineering, Mobile phones.

1 Introduction

Accompanying with the escalation of living standard, users' expectations on surrounding products are raising as well. Functionality alone can no longer satisfy user's demands. How to increase emotional value [2] of a product plays an important role in today's business strategy. Hence, Kansei Engineering (KE), employing engineering approach, was developed to find out which design characteristics elicit particular subjective feelings from people, and then build them into a product to elicit the desired responses [10, 17, 22]. Various studies have proved the usefulness of Kansei engineering, especially in the area of visual Kansei studies [14, 17]. However, without accurate measuring by scientific instruments, the results of such studies often pointed to a vague set of product elements instead of more specific ones for evoking certain Kansei feelings.

On the other hand, various researches conducted in different regions showed different results. It indicated that different cultural origins with different traditions, custom, ethic and values may have contributed to the different findings. Apparent evidences include Italian improvisatory and romantic flair, German precise and systematic orderliness, and American innovative and rich varieties, Japanese delicate and ethereal details, and French noble and fashionable touch [24].

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 47–56, 2007.
© Springer-Verlag Berlin Heidelberg 2007

1.1 Purposes and Objectives

Therefore, the purposes of this study were set to explore the cross cultural perspectives toward Kansei design using mobile phones as examples. Formal features, Kansei adjectives and the relationships between them were investigated with Kansei Engineering procedures. Three objectives can be drawn as follows: (1) exploring users experience and the preferences of mobile phones among different cultural backgrounds; (2) investigating the cultural effects on users' preferences of mobile phone designs and Kansei images; and (3) generating a set of culture dimensions for Kansei design.

1.2 Processes and Steps

Interviews and questionnaires are conducted along with the procedures of Kansei evaluations which can be divided into five steps: (1) collecting mobile phone samples and Kansei words; (2) selecting representative mobile phone samples and Kansei words via interviewing with experienced designers and KJ method respectively; (3) extracting design elements; (4) designing four sets of bilingual questionnaires with 5-point Licker Scale; (5) conducting experiments on four sites with questionnaires; and (6) analyzing results using Quantification Type I.

1.3 Subjects and Scopes

Top selling mobile phones from Taiwan, China, Japan and Korea are used as samples for this study. Brands include: Alcatel, Asus, BenQ, BenQ-Siemens, DoCoMo, Dopod, LG, Motorola, Nokia, OKWAP, Panasonic, Pantech, Samsung, Sharp, Sony-Ericsson, Toshiba, and etc.

2 Theoretical Bases

Related researches and literatures including cultural studies, formal features, and Kansei engineering are reviewed in order below to form the foundations for this study.

2.1 Cultural Studies

Hofstede [12] has conducted a cultural study on IBM's staffs from 64 countries during 1978 and 1983 and found that the differences were from different values. He then constructed them into five so-called dimensions of culture. They are: power distance, individualism vs. collectivism, masculinity vs. femininity, uncertainty avoidance, and long-term orientation. Hofstede's study [13] shown in Table 1 depicts that, compared to other three regions, Japan has the highest tendencies on individualism, masculinity and uncertainty avoidance, and the lowest on power distance; South Korea the lowest tendencies on masculinity and long-term orientation; Taiwan the lowest tendency on individualism; while China having the highest tendencies on power distance and long-term orientation, and the lowest on uncertainty avoidance. On the other hand, Barber and Badre [1], in their study of cultural characteristics of interface design, identified a

set of frequently used and preferred interface design elements and labeled them as Culture Markers. In a study on color and culture among China, Japan, Egypt, France and United States, Boor and Russo [3] found that there were different meanings and reactions toward colors such as: red, blue, green, yellow and white among them.

Mobile phones have been shaped under different culture and, in turn, influenced the cultural settings surrounding them. For example, in Japan, Keitai, short for mobile phones, was designed as compact as possible to meet Japanese carrying needs. And in turn, it has changed the commuting culture from reading news paper, magazine or book into pressing buttons on a Keitai [12]. In Korea, that more manufacturers has formed a severe competition to rapidly react to the market demands and to offer various designs have made Korean replace their mobile more frequently than any other countries. The only similarity between Japanese and Korean is the slogan: "Everything over mobile". While the development of mobile communication is getting matured, the talking time also gets reduced in most countries except Taiwan [5]. Even though the functions on a mobile have increased to an uncountable number, to Taiwanese conversation is still the most frequently used one. To China, mobile is not only a communication devise but also a sign of grown-up. Like in Japan, the mobile has developed a so-called message culture or thumb culture in China due to the system rate policy and their fond of sending short messages in daily life.

Table 1. Hofstede's dimension of culture scales. (http://www.kwintessential.co.uk/intercultural/dimensions.html)

Country	Power Distance	Individualism	Uncertainty Avoidance	Masculinity	Long-term Orientation
Japan	54	46	92	95	80
South Korea	60	18	85	39	75
Taiwan	58	17	69	45	87
China	80	20	30	66	118

2.2 Formal Features

Products communicate stylistic messages via various forms and features. Chen and Owen [6] proposed a Style Description Framework (SDF) equipping designers with the abilities to analyze existing styles and to describe new styles for target markets. As the product of a SDF, a "style profile" consists of a set of polar adjective scales and associated weighting mechanisms. Within the profile, stylistic attributes -- in the form of values given on the scales -- are grouped into six categories: form elements, joining relationships, detail treatments, materials, color treatments and textures. Two weighting mechanisms, an importance index and a confidence factor, fine tune the description. The "style profile" can be used not only to communicate styles between designers and computers but also to accumulate formal style knowledge.

Following the rules of Gestalt psychology and product aesthetic, Wallace [25], in his thesis, proposed a computer system capable of generating forms conforming to both aesthetic and manufacturing requirements. It can be summarized into four parts: (1) arrangement of the components according to aesthetic rules; (2) definition of the styles according to the types of edges and corners (eg: Braun style, High-tech style or

Art Deco style); (3) configuration of the product modules and components; and (4) graphics, textures and color treatments of the product surface.

Breemen [4] classified the aesthetic characteristics of product forms into three levels according to their contributions. Among them, form detail, constructing method, color, material, texture and light have the most contribution to product aesthetics. Product overall shape comes the second while the geometric space coordinate positioning having the least effect on product aesthetics.

2.3 Kansei Engineering

Emotional consuming is becoming a trend in global market competition. Consumers pay more and more attention on personal emotional feelings while buying things, which makes consuming-style change significantly [9]. Kansei Engineering (KE), a consumer-oriented new product development technique shaped to meet such trend, emphasizes the exploration of relationships between people's emotional feelings and artifacts' characteristics [7, 18, 20, 21]. Hence, it has become an important topic for user-oriented product development and design and a key factor in elevating design competence. Recent studies in this area have accumulated fruitful findings and demonstrated its value in product positioning during new product development and design stage [15, 19]. However, most of the studies focused on the mappings between single Kansei word and design elements [11], and left multi-Kansei evaluation intact.

In general, there are five steps in a Kansei Engineering process. (1) Selecting Kansei words - factor analysis or KJ method can be employed. (2) Selecting representative samples – those demonstrating well the Kansei words decided above are to be selected as samples. (3) Extracting essential form characteristics - experienced product designers can be called up to help extract most prominent design components (equivalent to items in KE) which contribute best to the Kansei words decided above. Possible design options (equivalent to categories in KE) can be further set for these components. (4) Constructing 3D digital samples - product pictures are generated for later Kansei evaluation. (5) Kansei evaluation - subjects are asked to evaluate the Kansei words against the product pictures using either Likert scale or Semantic Differential scale. Data collected are then analyzed with Quantitative Theory Type I to establish the relationships between each Kansei word and design elements.

3 Processes

The study was conducted with the following steps : (1) collecting samples and Kansei words; (2) selecting representative samples and Kansei words; (3) extracting design elements; (4) designing questionnaires; (5) conducting experiments with questionnaire; and (6) analyzing results using Quantification Type I.

3.1 Collecting Samples and Kansei Words

Mobile phone models marketed in these four nations are extensively collected, roughly 100 pieces. Four criteria used for screening the models are: (1) design paradigm – models chosen should be generally acknowledged as design paradigm with high value, as well as high selling volume and broadly discussed; (2) culture breadth – models

chosen should cover all of the four regions and be highly regarded; (3) style range – models chosen should cover a range of various styles; and (4) progressive trend – models chosen should be able to exhibit the progressive trend of mobile phone design within 2005 and 2006. Figure 1 shows some examples of them. Brands include: Alcatel, Arcoa, Asus, BenQ, BenQ-Siemens, DoCoMo, Dopod, Eten, Gigabyte, KDDI, LG, Motorola, Nokia, OKWAP, Panasonic, Pantech, Samsung, SCH, Sharp, Sony-Ericsson, Toshiba and etc.



Fig. 1. Mobile phones currently selling in Japan, South Korea, Taiwan and China (partial)

The most popular Kansei words (top 20 per country) used in describing mobile phones are collected from web pages and magazines published in theses four nations (Table 2). The top word of each cell appears in the original language while the bottom one the translation. For example, Chinese (traditional, simplified or Japanese Kanji) are the original language for all regions except South Korea. In other words, Chinese is the translation to South Korea while English the translation to the rest of the regions.

3.2 Selecting Representative Samples and Kansei Words

Representative mobile phones were selected after interviews with experienced mobile phone designers and users to cover all of the features usually exhibited on it. Figure 2 shows part of the chosen samples.

Interviews were designed into two versions: (1) in-depth interview for gathering knowledge and viewpoints from design experts, and (2) contextual exploration for gathering consumption and use experience from experienced users. The leading questions for design experts contain: personal experiences in mobile phone design, design strategy and corporate image of the belonging company (or studio), and design approaches applied against culture issues, while that for experienced users containing: personal experiences in owning and using mobile phones, and culture cognition of the mobile phones.

Table 2. Kansei words collected from Japan, South Korea, Taiwan and China

Japan	South Korea	Taiwan	China
Characteristic	Innovative	Technical	码的 Digital
Functional	Thin	Fashionable	的 Handy
快適 Cozy	Compact	味 Taste	的 Top Class
Minimal	Fashionable	Terse	的 Exquisite
Charming	Stylish 格	Delicate	潮流 的 Trend Pioneering
利 Handy	Colorful 艷	Durable	性的 Particular
Particular	Handy 利	Fresh	薄的 Flimsy
可愛 Cute	Unique	Fashionable	流行的 Fashionable
Flimsy	Functional	Tailored	的 Eye-Catching
Modern	Smart	Unique	的 Not Hackneyed
Advanced	High Tech	Particular	的 Intelligent
Unique	Curvaceous	Compact	的 Classic
單 Simple	Smooth 暢	巧 Light	的 Fine
安 Relieved	Palmary	可愛 Cute	内 的 Restrained
Atmospheric	Shining	經 Classic	的 Luxurious
Artistic	Cute 可愛	Thin	优 的 Elegant
麗 Gorgeous	Charming	級 High Class	爱的 Cute
表 Expressive	Sporty 動	年 Young	另 的 Out of Character
級 High Class	Crazy	商 Commercial	的 Dynamic
Linear	Magic 神	Leisure	酷派的 Cool

Each interview lasts about 80 minutes, including: introduction and camera setting (10 min.), interview (30 min.), experiment (30 min.) applying Evaluation Grid Method [23], following up questions (5 min.) and summing up (5 min.). Owing to the limited budget, interviews were only conducted in Taiwan.

KJ method was employed to group the representative Kansei words. Those ones which share higher popularity among different regions are chosen. Table 3 shows the result of the process.



Fig. 2. Representative mobile phones chosen for the Kansei evaluation (partial)

3.3 Extracting Design Elements

A list of design components (equivalent to items in KE) and design options (equivalent to categories in KE) was compiled after a series of in-depth interviews with experiment applying Evaluation Grid Method conducted for design experts. Table 4 shows the results.

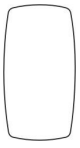
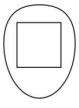
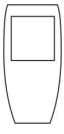
3.4 Designing Questionnaires

Before conducting the experiment on site, four sets of bilingual questionnaires were designed. Participants included design students and consumers in the mobile phone stores. The contents of the questionnaire consist of two parts: Kansei related and preferences related. Likert scale of 5 levels was used for the scoring.

Table 3. Representative Kansei words chosen for latter use

Japan	South Korea	Taiwan	China
可愛 Cute	Cute 可愛	可愛 Cute	爱的 Cute
 Flimsy	Thin	 Thin	薄的 Flimsy
 Modern	Fashionable	 Fashionable	潮流 的 Trend Pioneering
利 Handy	Handy 利	 Durable	的 Handy
 Functional	Functional	 Technical	码的 Digital
	High Tech	 Fresh	
	Innovative		

Table 4. Design Elements (Items and Categories)

Items	Categories			
A Body	 a1.	 a2.	 a3.	 a4.
B Screen	 b1.	 b2.	 b3.	
C Panel	 c1.	 c2.	 c3.	
D Buttons	 d1.	 d2.	 d3.	

3.5 Conducting Experiments

Questionnaires are then used with the Kansei evaluation experiments. Results are shown and analyzed in the following sections.

3.6 Analyzing Results

From Table 3, we discovered that some Kansei words share high similarity among the four regions while some are uniquely used for a single region. The former include: Cute, Thin (Flimsy), Fashionable (Modern or Trend Pioneering), Handy, Functional, and etc. while the later covering Expressive, Sporty, Young, Out of Character, Crazy, Commercial, Dynamic, Linear, Magic, Leisure, Cool, and etc.

From Table 4, the design elements extracted from design experts show that there are four major parts of a mobile phone contributing to Kansei image. They are: body, screen, panel and buttons.

4 Conclusions and Discussions

The Kansei needs of consumers from different culture backgrounds are different due to their diverseness in use habit. For example, in Japan, Keitai, short for mobile phones, was designed as compact as possible to meet Japanese carrying needs. And in turn, it has changed the commuting culture from reading news paper, magazine or book into pressing buttons on a Keitai. In Korea, that more manufacturers has formed a severe competition to rapidly react to the market demands and to offer various designs have made Korean replace their mobile more frequently than any other countries.

The preferred formal features of a mobile phone among different cultural backgrounds are different too. The only similarity between Japanese and Korean is the

slogan: “Everything over mobile”. While the development of mobile communication is getting matured, the talking time also gets reduced in most countries except Taiwan. Even though the functions on a mobile have increased to an uncountable number, to Taiwanese conversation is still the most frequently used one. To China, mobile is not only a communication devise but also a sign of grown-up. Like in Japan, the mobile has developed a so-called message culture or thumb culture in China due to the system rate policy and their fond of sending short messages in daily life. The results can be used as reference for designing cross-culture mobile phones as well as other closely related products.

Acknowledgments. Thanks to the financial support of National Science Council, Taiwan for this research under the contract number: NSC-95-2221-E-006-141.

References

1. Barber, W., Badre, A.: Culturability: The Merging of Culture and Usability. In: Proceedings of the 4th Conference on Human Factors and the Web, Basking Ridge, New Jersey (1998)
2. Barlow, J., et al.: Emotional Value: Creating Strong Bonds With Your Customers, Berrett-Koehler Pub (2000)
3. Boor, S., Russo, P.: How Fluent Is Your Interface? Designing for International Users, INTERCHI ’93 (1993)
4. van Breemen, E.J.J., Slamet, S.: The Role of Shape in Communicating Designers’ Aesthetic Intents. In: Proceeding of the 1999 ASME Design Engineering Technical Conferences, Las vegas, Nevada (1999)
5. Chen, H.P.: 2003. Mobile Cultural Revolution. Taipei: E-earthgeo (2003)
6. Chen, K., Owen, C.L.: Form Language and Style Description. Design Studies 18, 249–274 (1997)
7. Chen, K., Shing-Sheng G., Yi-Shin D., Yu-Ming C.: A Method for Converting Sensibility into Sense. Industrial Design, vol 29(1), pp. 2–16, Ming-Chi Institute of Technology, Taiwan (2000)
8. Cleveland, D., Cleveland, N.: Eyegaze eyetracking system. Imagina: Images Beyond Imagination. Eleventh Monte-Carlo International Forum on New Images, LC Technologies, Inc. (1992)
9. Gobe, M., et al.: Emotional Branding: The New Paradigm for Connecting Brands to People, Allworth Press (2001)
10. Harada, A.: The Parallel Design Methodology in the KANSEI Engineering. Report of Modeling the Evaluation Structure of Kansei, pp. 309–316 (1998)
11. Heo, S., Harada, A.: Research on Characteristics of Kansei Reaction toward Images, 4th Asian Design Conference, Japan (1999)
12. Hofstede, G.: Cultures and organizations: Software of the mind. McGraw-Hill, London (1997)
13. http://www.kwintessential.co.uk/intercultural/dimensions.html
14. Jindo, T., Hirasago, K., Nagamachi, M.: Development of design Support system for office chairs using 3-D graphics. International Journal of Industrial Ergonomics 15, 49–62 (1995)
15. Lee, E., Lee, H., Kim, M.: The Effects of Visual and Auditory Information as a Tool of Emotional Value Assessment, 4th Asian Design Conference, Japan (1999)

16. Liu, L.E.: Tokyo-Myrtle-Myth - Er's Japanese Affection. Taipei: Wheat Fields Publisher (2002)
17. Matsubara, Y., Nagamachi, M.: Hybrid Kansei Engineering System and design support. 1997) 19, 81–92 (1997)
18. McDonagh, D.: Visual Product Evaluation: Exploring Users Emotional Relationships with Products. Applied Ergonomics 33(3), 231–240 (2002)
19. Miyazaki, K., Matsubara, Y., Nagamachi, M.: A Modeling of Design Recognition in Kansei Engineering. Japanese Journal of Ergonomics 29(Special), 196–197 (1993)
20. Nagamachi, M.: Kansei Engineering. Tokyo: Kaibundo, Japan (1989)
21. Nagamachi, M.: Kansei Engineering: A consumer-oriented technology. In: Bradley, Hendrick (eds.) Human Factors in Organizational Design and Management - IV, pp. 467–472. Elsevier Science, Amsterdam (1994)
22. Nagamachi, M.: Kansei Engineering: A new ergonomic consumer-oriented technology for product development. International Journal of Industrial Ergonomics 15, 3–11 (1995)
23. Sanui, J.: Visualization of users' requirements: Introduction of the Evaluation Grid Method. Proceedings of the 3rd Design & Decision Support Systems in Architecture & Urban Planning Conference 1, 365–374 (1996)
24. Tsai, Z.W.: A Study on Product Image Language - Using Native Image as Example. Master Thesis, National Cheng Kung University, Tainan, Taiwan (1994)
25. Wallace, D.D.: A Computer Model of Aesthetic Product Design: an Approach. Master Thesis of Science in Mechanical Engineering at MIT, pp. 42–44 (1991)

The Challenge of Dealing with Cultural Differences in Industrial Design in Emerging Countries: Latin-American Case Studies

Alvaro Enrique Diaz

Lecturer Université de Montréal, invited lecturer El Bosque University, Colombia
473 Lusignan, Montréal, Qc, H3C 1Y7, Canada
alvaro.diaz@umontreal.ca

Abstract. Recent trends in industrial design for emerging markets have focused on the economies of China, India and some countries of Latin America. Even though those countries have opened up their markets (and their economies have grown rapidly during the past decade), companies still struggle to get reliable information about their domestic consumers. Foreign manufacturers try to understand local markets to find major opportunities for new investments, and therefore, specialists in marketing and human factors are required to find innovative strategies to deal with cultural differences. In many cases, products and services need to be redesigned for these new markets. Three case studies in Latin-America (Mexico, Colombia and Nicaragua) - in which ethnographic research was required to understand users' needs - exemplify this process.

Keywords: Industrial design, Human factors, Cultural design, Ethnographic studies, Usability evaluation, Latin America.

1 Introduction

Industrial design products for emerging economies in Latin America have called the attention of US and European manufacturers. The large market and the buying power of the population appear attractive to many of these companies[1]. Manufacturers must develop innovative strategies to understand the particularities of these environments and cultural differences that can become opportunities to develop new products and services. Even though the interest of some companies is to penetrate these large markets and to introduce customized services and products, they often struggle to get reliable and complete information about consumers in emerging countries [1].

1.1 Methodology

For this paper, we choose three projects in different areas, including two consumer products and a health care product. Fig. 1 represents some of the methods used to

[1] "Since 2004, Latin America's economy (GDP) growth as 5% in average", Reid Michael, "The ride ahead", The Economist, The world in 2007, 21st Edition, December 2006 [2].

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 57–64, 2007.
© Springer-Verlag Berlin Heidelberg 2007

achieve the objectives in each of the projects. This is not a linear process but instead a cyclic and holistic research approach. This method allows us to describe and understand the activities and needs of the population. In fact, in the ethnographic research multiple factors are judiciously considered:

- Customized protocols: Depending on the project's objectives and its environment, protocols should be customized. For instance, in the protocols used in consumer products the emphasis must be placed in the relation between the users and their needs. In other cases, such as the health care industry, protocols could be more oriented towards the relation between the patients and their illnesses.
- Interviews on site and observation: No matter if the research is done locally or in a foreign country, this approach reminds us of the importance of taking into consideration the influence of the cultural, social, political, and economic environment.
- Autoconfrontation method: Once some data has been collected through the protocols and the interviews, it's important to share some of the results with a selected group of users. Often, this can contribute to validate some the hypotheses with them.
- Cultural probes: It is important to identify relevant cultural factors, which can be used to generate a particular concept in each context.

Following this analysis, it is possible to understand – and represent - the environment; a brainstorming can help to conduct this exercise. Finally, Fig. 1 shows that all the information and ideas can be filtered in the form of a funnel in order to synthesize the key elements that will be later used for the design concepts.

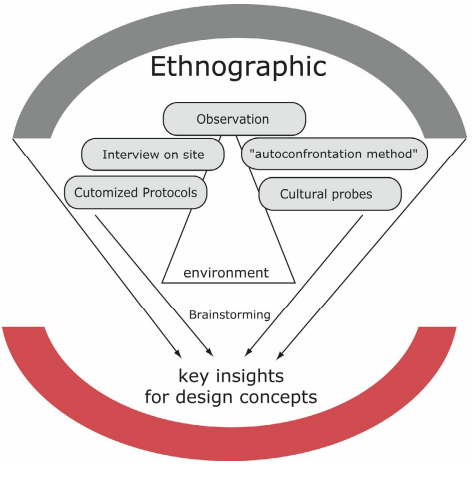


Fig. 1. Diagram of the methods used in the research – The ethnographic approach

2 Cases Studies

The projects exposed in this paper, were conducted in Nicaragua, Mexico and Colombia. In Nicaragua, a health product was tested and a prototype was used to

conduct the research. In Mexico, in which a research group tried to find innovative design opportunities, previous remote research provided data to compare with an empirical study. The Colombian project was conducted in order to identify users' needs in a specifically targeted product.

2.1 Case Study 1: Prostheses Leg for Developing Countries

The main objective of this project was to propose a tool to assist people with reduced mobility (PWRM) in developing countries by designing a prosthesis leg particularly adapted for this market. Reduced mobility limits the opportunities to have a job and to be integrated in the society. In developing countries, this is an even more difficult challenge due to the high prices of high-tech prostheses.

Nicaragua was chosen for the study[2]; on a first step we collected information about existing prostheses around the world[3]. We found that the energy prostheses concept (Fig. 2) could be an appropriate solution for this situation. We found that this solution responds better to PWRM due to the advantages obtained from reducing the effort to walk and from increasing users' capacity to run.



Fig. 2. Advanced energy-storing prostheses[4]

Unfortunately, these prostheses are not affordable for low income people in Nicaragua (prices of this prostheses range between 1300$US and 3800$US per unit). A first prototype particularly adapted to the Nicaraguan context was made by the designer Sébastien Dubois and it was later tested in the region. Protocols were also conducted to interview local patients, nurses and doctors. Methods for collecting data included photographic and video documentation and surveys about users' lifestyle and activity goals. In 2006, an affordable prototype that costs 10$US was produced with a local manufacturer. The prototype was tested (Fig. 3) with the following results:

---

[2] After Hurricane Micth, researchers struggled to find mines that were moved by rain, water and floods. According to Mine Action Information Center at James Madison University, "In 2000 the army suggested that there is one mine in the ground for every 55 Nicaraguans."

[3] This project was lead by designer Sébastien Dubois and was supervised by designer Alvaro Enrique Diaz.

[4] Brian J. Hafner, BS; Joan E. Sanders, PhD; Joseph M. Czerniecki, MD; John Fergason, CPO, Transtibial energy-storage-and-return prosthetic devices: A review of energy concepts and a proposed nomenclature Journal of Rehabilitation Research and Development Vol. 39 No. 1, January/February 2002 Pages 1-11 [3].

1. The idea of designing "conservative" prostheses (not fashionable at all) was quickly ruled out by the designer once he found out that Nicaragua's standards were more related to "Modern, high-tech and styling" prostheses like the ones that can be often found in developed countries.
2. An affordable prototype was produced due to the manufacturing process established in Nicaraguan labs.
3. When developing the concept, humidity needs to be taken into consideration, due to the geographical location, as well as the activities related with the use of the prosthesis.
4. Prostheses need to be used in irregular lands.

As a consequence, the new design needs to have a modern style and the materials used, to produce the prosthesis, can be found locally. The concept is still in development and through the support of Handicap International more tests will be done this present year.



Fig. 3. Test in Nicaragua[5]

2.2 Case Study 2: Mobile Phones in Latin America

In some Latin American countries, the mobile phone industry is composed of numerous informal businesses. In many Colombian cities, for example, low income residents informally "sell minutes" as an income generation activity. Such as it was common to see people selling juices, ice cream, fruits in the middle of the street, we now have these same people selling as well "minutes" and several accessories used for cell phones, and all this without any legal permission.

In Early 2006, there were around 241 million mobile phones in Latin America[6]. Some specialists predict that over 50% of Latin Americans will own a mobile phone by 2007. In 2005 an independent American nonprofit corporation involved internationally

[5] Dubois Sébastien, Qualité de vie des personnes handicapées vivant dans les pays en voie de développement, Unpublished thesis research for obtaining a Bachelor degree in Industrial Design, University of Montréal, Montréal April 2006 [4].

[6] Paul Budde Communication Pty Ltd., May 2006, Pages: 710 Telecoms, Mobile and Broadband in Latin America - 2006 – Geographic, online on Dec.15th-2005, http://www.researchandmarkets.com/reports/c45331, in Forbes Magazine, November 15th 2006 [5].

with scientific research and technology development projects, organized an immersion workshop to find product and business opportunities in the emerging segment of the Mexican market [6]. A multinational consortium represented by companies of the pharmaceutical industry, the health care industry and other manufacturers of consumer products, sponsored this project to find out the characteristic of this market.

The first meetings were made domestically in an American city. The research group chose some neighborhoods with a big influence of Mexican immigrants. In a second phase, the research group moved to Michoacán, Mexico, to validate the data.

Some design recommendations were made through 39 design concepts proposed for future innovation initiatives. Some of the recommendations were:

- To consider the importance of the family, culture and community loyalty in all proposed concepts.
- To account for the distinction between documented immigrants, undocumented immigrants and US born citizens.
- To include informal business practices.

These recommendations were used to generate design concepts including:

- Multi-purpose retail locations and modular kiosks.
- Developing industry in the communities as an OEM (original equipment manufacturer) for local product assembly.

2.3 Case Study 3: A Stove for the Base of the Pyramid[7]

This project was started in Ciudad Bolivar, one of the poorest neighborhoods in Bogota, Colombia. The main objective was to develop a new design for a "20$ stove". The environment of the users, their motivations and needs were the main focus of the study.

Even though the main objective of the research was to develop a new stove, interesting points were found about the "way" to cook. During the research, we used some methods to quantify manufacturing considerations as well as cooking considerations and some questionnaires to understand users' opinions about the project. Fig.4 shows some manufacturing considerations that had to be considered. In this case, the number of welding joints was identified as an important consideration due to the limited budget that users have for a new stove. We compared the new design with three previous concepts found on the market. Critical considerations were marked in red and interesting considerations were marked in orange.

In the Fig.5, we can see functional considerations as the total area to use in each grill.

With the functional considerations we also identified the basic needs of this population. Users constantly complaint about their economic situation, "we don't have enough money to buy a four burner gas stove, even if we can have some savings it will be impossible to afford the price of the gas". Questionnaires helped us to

[7] Diaz Alvaro Enrique, "Stove for basic needs", Design Research Project, Bogotá, Colombia 1997.

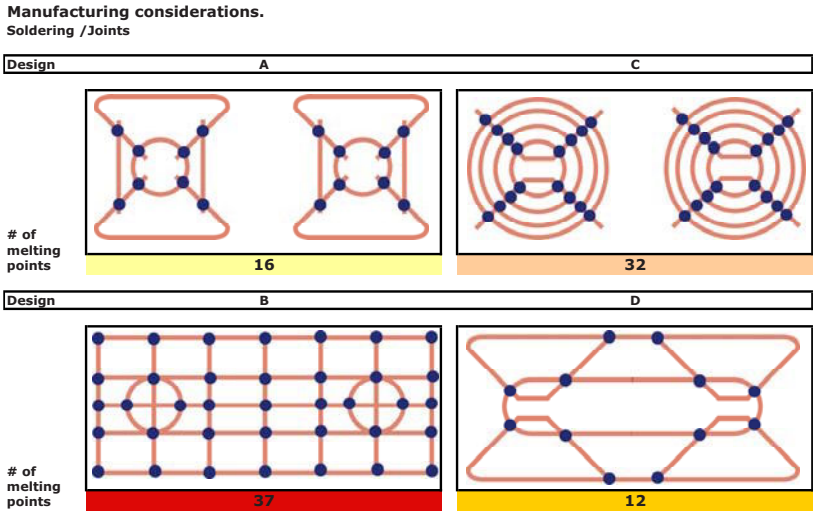


Fig. 4. Manufacturing considerations

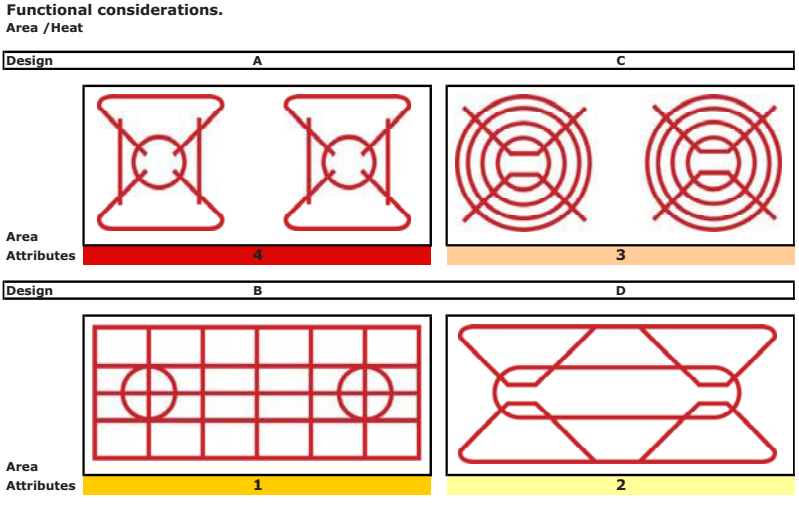


Fig. 5. Functional considerations

understand their diet and to identify their discomfort with the "two burner gas stove".Meat, rice, soup and salad were the daily basic diet, so when they were cooking, they had, for example, to take away the soup to have space to cook the meat. Fig.6 shows that during this process, the soup got cold. Their habits showed us that they used diner time as an important moment of their journey to eat and talk with relatives.

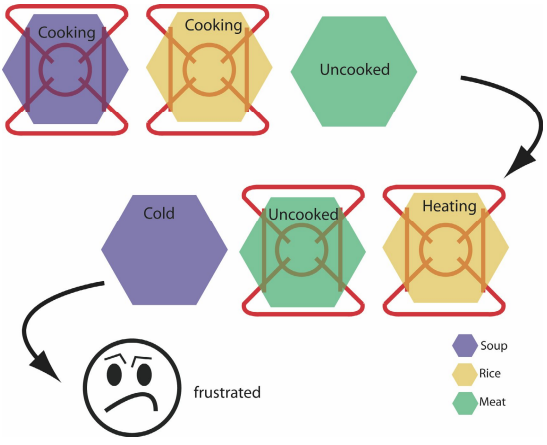


Fig. 6. Regular process

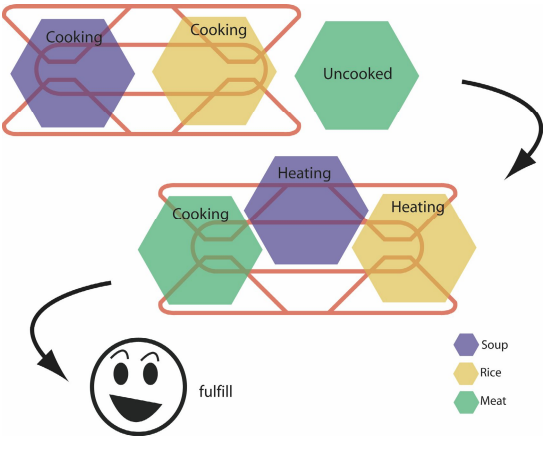


Fig. 7. Innovative process

The most important issues discovered in this research were:

- More than an aesthetic problem, the basic needs were to cook the food with the minimum of gas.
- Data gathering showed that the diet was very similar than the one of other socio-economic levels in Colombia.
- Important meals were soup, rice and meat.

The recommendations were to simplify and improve the cooking process. Opportunities of design were focused on the grill.

3 Conclusion

In the three cases, data collection was guided by detailed protocols designed to focus the investigations and to facilitate the understanding of social and cultural trends and product use in emerging countries. Without insight input from these markets, designing culturally, socially and economically relevant products would have been impossible for the designers.

A sensitive approach is important to include cultural knowledge necessary to develop a thorough understanding of the intended users.

References

1. Khanna, T., Palepu, K.G., Sinha, J.: Strategies That Fit Emerging Markets. Harvard Business Review, vol. 83(6) (June 2005)
2. Michael, R.: The ride ahead. The Economist, The world in 2007, 21st edn. (December 2006)
3. Hafner, B.J., Sanders, J.E., Czerniecki, J.M., Fergason, J.: Transtibial energy-storage-and-return prosthetic devices: A review of energy concepts and a proposed nomenclature Journal of Rehabilitation Research and Development. vol. 39(1), pp. 1-11 (January/February 2002)
4. Sébastien, D.: Qualité de vie des personnes handicapées vivant dans les pays en voie de développement, Unpublished thesis research for obtaining a Bachelor degree in Industrial Design, Université de Montréal, Montréal (April 2006)
5. Paul Budde Communication Pty Ltd. May 2006, Pages: 710 Telecoms, Mobile and Broadband in Latin America - 2006 – Geographic, online on Dec.15th-2005, in Forbes Magazine (November 15th 2006) http://www.researchandmarkets.com/reports/c45331
6. Linda, P.: Design Research in Remote and Emerging Markets: Mexico, China, Thailand and India. Paper presented at the MX Design Conference, Iberoamericana University, Mexico City (2005)

Emerging Issues in Doing Cross-Cultural Research in Multicultural and Multilingual Societies

Henry Been-Lirn Duh[1] and Vivian Hsueh-Hua Chen[2]

[1] 50 Nanyang Avenue, Center for Human Factors and Ergonomics
Nanyang Technological University, Singapore, 639798
mblduh@ntu.edu.sg
[2] 31 Nanyang Link, Wee Kim Wee School of Communication and Information
Nanyang Technological University, Singapore, 637718
chenhh@ntu.edu.sg

Abstract. Cross-cultural research is one of the emerging areas in HCI field lately. There have been fruitful discussions on issues of using measurement or doing field work to address HCI issues 'across countries' or 'across-cultures' . However, methodological concerns in conducting research in multicultural and multilingual society have not been fully explored. This paper reviews research work done and outlines problems and concerns in doing cross-cultural research in multicultural and multilingual society/country. Consequently, we propose a conceptual framework/procedure as a starting point for further development of measurements or field strategies.

Keywords: Cross-cultural measurement.

1 Introduction

Cross-cultural research is one of the emerging areas in HCI field lately. During the 1990's cross-cultural HCI research has expanded from issuing guidelines and importing models from the social sciences [23] to developing its own frameworks [18]. For any company to design and market products for sales to other countries, it is critical to understand users' cultural background and possible consequential responses to products. Doing user studies across different cultures and countries becomes a common practice for researchers and practitioners. There have been fruitful discussions on issues of using measurement or doing field work to address HCI issues 'across countries' or 'across-cultures'. However, it becomes evident that an effective research tools used in one culture may not gather the data one wants in another culture. Methodological concerns in conducting research in multicultural and multilingual society have not been fully explored. This paper reviews research work done and outlines problems and concerns in doing cross-cultural research.

2 Research Methodology in Cross-Cultural Research

The nature of cross-cultural research is to understand and compare various cultural practices. It increases the complexity of research process because the researcher has to

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 65–73, 2007.
© Springer-Verlag Berlin Heidelberg 2007

consider various cultural factors that can contribute to the differences in research outcomes. The major challenge in doing research across cultures is to establish equivalence in each stage of the research process. No research is completely independent from the cultural context of the researcher. In order to obtain the data sought, the research tool and research procedure have to be suitable for the cultural context the research is conducted.

There are works done in multiple fields to enhance understanding of the issue. Majority of the literature has raised issue of research approach. The earliest work can be found in Frijda & Jahoda's [19] discussion on the difficulties to achieve equivalence in doing cross-cultural research in psychology. Many studies have then discussed equivalence in cross-cultural research in various ways. Sekaran [39] delineates methodological issues into five dimensions: functional equivalence, instrumentation, data-collection, sampling design and data analysis. Adler [2]adds criterion definition and research administration to Sekaran's [39] structure. Hui and Triandis [25] suggested that cross-cultural comparability can be achieved by establishing compatibility across cultures on four key categories of equivalence. Conceptual or functional equivalence, Construct operationalization equivalence, Item equivalence, Scale Equivalence Methodological simplicity and level of analysis were later added to update this same framework [34]. Carvusgil and Das [10] came up with four categorizations: basic research design, sampling issue, instrumentation & data collection and data analysis. Usunier & Lee [43] summarize equivalence issues into conceptual, functional, translation, measure, sampling and data collection equivalence.

The above list of categorization is not exhaustive, but reflects common major themes in existing literature over the past five decades. High degree of similarities is observed in the ways methodological issues are discussed. Scholars who work in this area also provide possible explanations and solutions to problematic issues in conducting cross-cultural research. It is evident that researchers who are interested in conducting cross-cultural research need to be aware of and deal with those common equivalence problems. However, the problem becomes more complex when a research is to be conducted in multicultural and multilinguistic environment [44]. The threats to research measures increase exponentially in a cross-national study as the diversity encompassed in such a project is expanded [14]. Research has not extended discussion on equivalence in such a complex context. This paper does not attempt to revise existing categorization. Rather, it aims at situating equivalence issues in multicultural and multilingual societies. In the next section, issues central to conducting cross-cultural research are identified. Directions to advance understanding of methodological those issues in multicultural context are discussed.

2.1 Conceptual Equivalence

People from different cultures understand the world around them in their specific ways. Hence, the basic issue in cross-cultural research is to understand if meaning constructed from various cultures can reach "sameness". The sameness in the way people assign meaning to concepts and behaviors is conceptual equivalence. Byrne & Watkins [7] warns that it is difficult to achieve conceptual equivalance in cross-cultural research. He attributes this to "item bias", a condition whereby due to cultural differences, people set for themselves different criteria to judge concepts by. Even the

most basic of concepts, such as human motivation, can encounter conceptual inequality issues. Spini [41] examined a concept "hedonism" from a 10 value scale showed strong conceptual inequality.

To understand factors influence conception, several dimensions have been identified from past research: perception, values and attitudes, personality and gender.

Perception. Any research trying to find out people's opinions or reactions would have to deal with a subjective perception of each individual. Material presented to the research participants may not be perceived in the same way across cultures. Hong, Benet-Martinez, Chiu and Morris [24] found that exposure of Chinese or American material to Chinese-American bilinguals activated a different knowledge system and affected the way they responded in different ways.

Being aware of this, some "inter-national" studies that aims at testing differences have tried to differentiate the degree of variance across countries. They want to know how different people perceive things? Ewing, Caruana & Zinkhan [16] found that perception to advertising is different between the United States and countries outside it. However, differences are usually less when comparing countries outside the United States, even when there is a big geographical distance (e.g. UK and Singapore).

Moreover, perception is dynamic and that makes it hard to research equivalence [36]. Weber and Hsee [48] found that while attitudes to risk are relatively anchored, perception to risk however is highly volatile. Attitudes need to be researched separately since it influences research outcomes in different ways.

Values and Attitudes. Every culture has a distinct value system. Researchers tend to design research questions/hypothesis in the way they place values to the issue. One obvious example is that Western constructs do not always correspond to the rest of the world. Farruggia, Chen, Greenberger, Dmitrieva, and Macek [17] pointed out that their assumption of self esteem correlate to depression does not stand true in collectivistic society.

Personality. Personality is an important indicator for conceptual understanding. Leung and Bond [28] suggested that individual traits and cultural traits need to be distinguished in doing comparative studies.

Gender. Literature in this area is relatively fewer than the rest of the issues identified. It recognizes possible differences in how males and females perceive things [38, 47]. However, there is no compelling result that recommend how gender differences play a part in methodology. This is an area for further research.

The current approach to understand conceptual equivalence typically assumes cultural homogeneity within a nation. Few research addressed heterogeneity within a given culture. Aspinall [3] and Kim, Li and Ng [27] indicate ethnic differences can contribute to strong difference in conceptual equivalence. In a multicultural society, cultural groups within the same country or geographical area have distinct way to conceptualize things in life. Failure to consider differences among groups of different cultural background within a country can lead to inaccurate research outcomes.

2.2 Measurement Applicability

A large body of literature addressed problems of applying measurement across cultures. One of the often-cited problems in the interpretation of cross-cultural differences is the lack of comparability of testing methods [5]. Indeed, achieving such comparability can seem like a daunting task, considering that over 50 types of equivalence have been discussed in the literature (see Johnson, 1998).

For example, a well-known measurement Job Description Index (JDI) has been tested multiple times and proven to be sufficient in cross-cultural research. However, Wang and Russel [46] point out that the index fails to uncover unique traits of satisfaction that exist if one takes an "emic" perspective. To complicate the problem, Morland & Williams [31] noted that scales used in cross-cultural research can be an indication of "direction", e.g. change in attitudes, etc. as opposed to a test of difference.

Another difficulty is the degree of generalizability. Paunonen and Ashton [35] reviewed a variety of personality measurements found that although the tests had "transportability", the ability to generalize findings from foreign cultures is limited. This substantiates that even though an instrument may have sufficient reliability, or measurement equality, generalizing is difficult in cross-cultural contexts.

It brings out the importance of addressing issues of contextual influence. A person of a certain culture may find dimensions of his culture manifesting in different degrees depending on the country he is located. If he is located within a foreign country, there is higher chance of manifesting the cultural aspects there in his answers. Chen, Ng & Rao [11] recommended to solve this problem by cultural priming. Cultural priming refers to removing cultural specific questions from measurements.

Another proposed solution for solving this problem is to examine "measure cultural similarities and differences at a more concrete level (i.e; specific, everyday attitudes and behaviors that seemed common to many societies at many points in time)" to avoid measurement issues. Funkhouser [20] calls these constructs "anchors", by which "bias-free, cultural comparisons can be made."

Studies that use non-verbal cues or pictorial representations as stimuli show various outcomes. Morris [32] writes that the self assessment manikin (SAM) is a highly efficient tool for measuring advertising responses, and that results show that this method of deriving emotional responses are "generally the same in the United States and Taiwan". On the other end of the spectrum, Hofer & Chasiotis [22] found the picture story method far too subjective. No consistency in that research was found. The article by Ye [49] also found that facial expressions are culturally bound and subjective. Facial expressions of Chinese varied with their own spoken linguistic cues.

2.3 Translation Quality

Translation quality refers to how similar the meaning of a term is to its original meaning after it is translated to another language. The "sameness" of the term can be subdivided into the following areas: direct vocabulary, semantic and grammar-syntactical. *Vocabulary equivalence* refers to the exact translation of the word. Does a term, love, for example, have a similar word in Japanese? *Semantic equivalence* notes the differences in meaning of a word. While love in English can be a generic expression for an emotion, there may be many forms and degrees of love in another

language. *Grammar-syntactical equivalence* is a degree of sameness in the construction of sentences. This is of more relevance to open ended responses, where individual translated words may have no meaning when placed as a sentence. Little studies is done in grammar-syntactical equivalence.

Rogler [37] expressed that translation inequality is a result of "cultural insensitivity". Not only does one have to know about the idiomatic or direct vocabulary similarity, knowledge of the country's history is also highly important. It is important for researchers, therefore, not only to get the hang of the language, but also to understand if the word can suit a country's history and experience.

The translational problems should not be taken just as the meanings of the question. Translation issues with regards to scales can also cause serious validity issues. Bad to excellent, as a scale for example, may not work well when the second language does not possess words that can accurately mean the above. This is particularly important since most of the scales and methods are developed in English.

After the translation is done, the next step is to test if the translation is of good quality. Several studies have dwelled on this issue [6, 8, 9, 42, 21]. One common suggestion is to back translate. For bilingual participants, asking them to answer both English and the translated terms is suggested. However, this method has also been criticized because bilingual respondents tend to have a "parent language" and hence think in a particular way that may be different from what a monolingual person might think. For a multilingual context, Duh and Save [15] have recommended to provide multiple languages in the measurement but allow participants to choose the language they are more comfortable with.

2.4 Response Styles

Baumbartner & Steenkamp [4] suggest that response style contamination has been overlooked in existing cross-cultural research. Not only do people not look out for response style issues, they also do not take up measures to prevent contamination (such as reverse coding). Response styles issues are divided into the following parts: extreme response styles, midpoint tendency, acquiesce response styles, and pleasing tendency.

Extreme response styles. Extreme response style is a tendency for the respondent to place his/her response on the extremities of a scale. Clarke [13] suggests that while it is easy to find out extreme response styles and sift it out, it is much more complicated when extreme response styles are mixed with acquiesce response styles. The study of cultures can reveal extreme response styles, and the researcher has to be prepared. Preemptive measures are recommended within the article. Hui and Triandis [26] also suggests complex reasons for extreme response styles. In their research, Hispanics were tested for extreme response styles. They demonstrated strong extreme response styles in a 5 point scale, but insignificant response style bias on a 10 point scale.

Some researchers have made suggestions. Arce-Ferrer and Ketterer [1] found that extreme response was slightly improved when items were added to moderate out the extreme response. Cheung and Rensvold [12] use Multiple-group confirmatory factor

analysis to test whether cultural groups can be meaningfully compared on the basis of factor (latent) means. Researchers can use this test to derive if their findings have any use after being contaminated by extreme response styles.

Midpoint tendency. Midpoint tendency is a response style where the respondent tends not to input extreme opinion using likert scales. Midpoint tendencies tend to be a problem within more conservative nations whereby expressions of strong opinions are not encouraged. Cheung and Rensvold [12] recommend avoiding this problem by using a "two stage format study".

Acquiesce Response styles. This is the classic "yea-saying" response styles. However, this form of response style can also be generalized as a strong one sided response style. When Mondak and Canache [30] removed the "neutral option" and introduced the "don't know option" in the questionnaire, it brought forth strong acquiesce response styles. People are unlikely to admit they do not know answers to the questions.

Researchers should also be more careful when they construct surveys or questions for cultures ranked high in uncertainty avoidance, as chances of acquiesce response styles emerging are higher [40]. Smith [40] writes that "Bias can be thought of as a nation-level reflection of the individual communication styles and patterns of inter-group relations that prevail within certain specifiable cultural contexts." He suggested adding in a factor that offsets the acquiescence bias.

Pleasing response styles. This is an interaction effect whereby the respondent tries to respond in a way to please the interviewer. It is suggested that developing countries tend to exhibit a social desirability bias. That causes problems especially when the researcher has no access to the real beliefs of an individual. When conducting interviews, researchers should also think about the race and gender of the interviewer and those interviewed, speculating if it might affect the interview in anyway.

2.5 Sampling

According to Lonner & Berry [29] "the problem of drawing equivalent samples will be a major, if not the most major, methodological obstacle to overcome satisfactorily." When doing cross cultural research, researchers will face a number of problems when it comes to their samples. One of them is "matching". "Matching" is described as getting 2 samples from 2 different cultures with (usually) identical socioeconomic statuses. Dorian, 2002 found it hard to find exact socio-economic samples for comparisons given that occupations, income levels, educational levels and spending parity are vastly different between China and North America. In that project, the researcher gave up the search for identical matches and instead looked towards a more qualitative, holistic approach to complete his research. Another article by Mullen, Milne & Doney [33] noted that outliers can be a threat to cross-cultural sampling. It can be harder to detect outliers in a cross-cultural setting.

3 Conclusion

This paper outlines important methodological concerns in doing user studies across different cultural contexts. Current research identified measurement applicability and translation quality as two primary issues when conducting cross-cultural research in a multicultural and multilinguistic context. However, very few research provided viable suggestions to solve the problems. Future research should explore potential problems in other areas as well as provide possible solutions to assure the degree of rigor in cross-cultural research.

References

1. Arce-Ferrer, A.J., Ketterer, J.J.: The effect of scale tailoring for cross-cultural application on scale reliability and construct validity. Educational and Psychological Measurement 63, 484–501 (2003)
2. Adler, N.: A Typology of Management Studies Involving Culture. Journal of International Business Studies 14, 29–47 (1983)
3. Aspinall, P.J.: Approaches to developing an improved cross-national understanding of concepts and terms relating to ethnicity and race. International Sociology 22, 41–70 (2007)
4. Baumbartner, H., Steenkamp, J.E.M.: Response styles in marketing research: A cross-national investigation. Journal of Marketing Research 38, 143–156 (2001)
5. Bond, M.H., Smith, P.B.: Cross-cultural social and organizational psychology. Annual Review of Psychology 47, 205–285 (1996)
6. Bontempo, R.: Translation Fidelity of Psychological Scales: An Item Response Theory Analysis of an Individualism-Collectivism Scale. Journal of Cross-cultural Psychology. 24, 149–166 (1993)
7. Byrne, B.M., Watkins, D.: The issue of measurement invariance revisited. Journal of Cross-cultural Psychology 34, 155–175 (2003)
8. Candell, G.L., Hulin, C.L.: Cross-language and cross-cultural comparisons in scale translations. Journal of Cross-cultural Psychology 17, 417–440 (1986)
9. Chapman, D.W., Carter, J.F.: Translation procedures for the cross cultural use of measurement instruments. Educational Evaluation and Policy Analysis 1, 71–76 (1979)
10. Carvusgil, S.T., Das, A.: Methodological issues in empirical cross cultural research: a survey of the management literature and a framework. vol. 37, pp. 71–96 (1997)
11. Chen, H., Ng, S., Rao, A.R.: Cultural differences in consumer impatience. Journal of Marketing Research 42, 291–301 (2005)
12. Cheung, G.W., Rensvold, R.B.: Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. Journal of Cross-Cultural Psychology 31, 187–212 (2000)
13. Clarke, I.: Extreme response style in cross-cultural research. International Marketing Review 18, 301–324 (2001)
14. Davis, H.L., Douglas, S.P., Silk, A.J.: Measure unreliability: A hidden threat to crossnational marketing research? Journal of Markering Spring 45, 98–108 (1981)
15. Duh, H.B.L., Reva, R.: Technical Report: usage of affective words in multilingual society. Center for Human Factors and Ergonomics, Nanyang Technological University, Singapore (2005)

16. Ewing, M.T., Caruana, A., Zinkhan, G.M.: On the cross-national generalizability and equivalence of advertising response scales developed in the USA. International Journal of Advertising 21, 323–343 (2002)
17. Farruggia, S.P., Chen, C., Greenberger, E., Dmitrieva, J., Macek, P.: Adolescent self-esteem in cross-cultural perspective. Journal of Cross-cultural Psychology 35, 719–733 (2004)
18. French, T., Smith, A.: Semiotically enhanced Web Interfaces for Shared Meanings: Can Semiotics Help Us Meet the Challenge of Cross-Cultural HCI Design? IWIPS 2000 Baltimore, US (2000)
19. Frijda, N., Jahoda, G.: On the scope and methods of cross-cultural research. International Journal of Psychology 1, 109–127 (1966)
20. Funkhouser, G.: A self-anchoring instrument and analytical procedure for reducing cultural bias in cross-cultural research. Journal of Social Psychology 133, 661–673 (1993)
21. Geisinger, K.F.: Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment information. Psychological Assessment 6, 304–312 (1994)
22. Hofer, J., Chasiotis, A.: Methodological considerations of applying a tat-type picture-story test in cross-cultural research. Journal of Cross-cultural Psychology 35, 224–241 (2004)
23. Hofstede, G.: Cultures and organizations: Software of the mind. McGraw-Hill, London, England (1991)
24. Hong, Y., Benet-Martinez, V., Chiu, C., Morris, M.W.: Boundaries of cultural influence: Construct activation as a mechanism for cultural differences in social perception. Journal of Cross-Cultural Psychology 34, 453–464 (2003)
25. Hui, C.H., Triandis, H.C.: Measurements in cross-cultural psychology. Journal of Cross-Cultural Psychology 16, 131–152 (1985)
26. Hui, C.H., Triandis, H.C.: Effects of culture and response format on extreme response style. Journal of Cross-Cultural Psychology 20, 296–309 (1989)
27. Kim, B.S.K., Li, L.C., Ng, G.F.: The Asian-American values scale-multidimensional: Development reliability and validity. Cultural Diversity and Ethnic Minority Psychology. 11, 187–201 (2005)
28. Leung, K., Bond, M.H.: On the empirical identification of dimensions for cross-cultural comparisons. Journal of Cross-Cultural Psychology 20, 133–151 (1989)
29. Lonner, W., Berry, J.: Sampling and surveying. Field methods in cross-cultural research Sage Publications, Inc., pp. 85–110 (1986)
30. Mondak, J.J., Canache, D.: Knowledge variables in cross-national social inquiry. Social Science Quarterly 85, 539–558 (2004)
31. Morland, J.K., Williams, J.E.: Cross-cultural measurement or racial and ethnic attitudes by the semantic differential. Social Forces 48, 107–112 (1969)
32. Morris, J.D.: Sam: The self assessment manikin. Journal of Advertising Research 35, 63–68 (1995)
33. Mullen, M.R., Milne, G.R., Doney, P.M.: An international marketing application of outlier analysis for structural equations: A methodological note. Journal of International Marketing 3, 45–62 (1995)
34. Nasif, E.G., Al-Daeaj, E.B., Thibodeaux, M.S.: Methodological Problems in Cross-cultural Research: An updated Review. Management International Review 31, 79–91 (1991)
35. Paunonen, S.V., Ashton, M.C.: The structured assessment of personality across cultures. Journal of Cross-Cultural Psychology 29, 150–170 (1998)

36. Ployhart, R.E., Wiechmann, D., Schmitt, N., Sacco, J.M., Rogg, K.: The cross-cultural equivalence of job performance ratings. Human Performance 16, 49–79 (2003)
37. Rogler, L.H.: Methodological sources of cultural insensitivity in mental health research. American Psychologist 54, 424–433 (1999)
38. Schwartz, S.H., Rubel, T.: Sex differences in value priorities: Cross-cultural and multimethod studies. Journal of Personality and Social Psychology 89(6), 1010–1028 (2005)
39. Sekaran, U.: Methodological and Theoretical Issues and Advancements in Cross-Cultural Research. Journal of International Business Studies. 14, 61–73 (1983)
40. Smith, P.B.: Acquiescent response bias as an aspect of cultural communication style. Journal of Cross-Cultural Psychology 35, 50–61 (2002)
41. Spini, D.: Measurement equivalence of 10 values types from the Schwartz value survey across 21 countries. Journal of Cross-cultural Psychology 34, 3–23 (2003)
42. Thomas, D.L., Weigart, A.J.: Determining nonequivalent measurement in cross-cultural family research. Journal of Marriage and the Family 34, 166–177 (1972)
43. Usunier, J-C., Lee, J.A.: Marking Across Cultures, 4th edn. Prentice Hall, London (2005)
44. Van der Vijver, F.J.R., Leung, K.: Methods and Data analysis for cross cultural research. Sage, Thousand Oaks, CA (1997)
45. Van Raaij, W.F.: Cross-cultural research methodology as a case of construct validity. Advances in Consumer Research 5, 693–701 (1978)
46. Wang, M., Russel, S.S.: Measurement equivalence of the job descriptive index across Chinese and American workers: Results from confirmatory factor analysis and item response theory. Educational and Psychological Measurement 65, 709–732 (2005)
47. Watkins, D., Cheung, S.: Culture, gender and response bias: An analysis of responses to the self-description questionnaire. Journal of Cross-Cultural Psychology 26, 490–504 (1995)
48. Weber, E.U., Hsee, C.: Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards perceived risks. Management Science 44, 1205–1217 (1998)
49. Ye, Z.: The chinese folk model of facial expressions: A linguistic perspective. Culture and Psychology 10, 195–222 (2004)

The Digital and the Divine: Taking a Ritual View of Communication and ICT Interaction

Brooke Foucault[1] and Jay Melican[2]

[1] Northwestern University, Center for Technology and Social Behavior, 2240 Campus Drive, Room 2-431, Evanston, IL 60208, USA
b-foucault@northwestern.edu
[2]Intel Corporation, Digital Home Group, Domestic Designs & Technologies Research, 20270 NW AmberGlen Court, AG1-112, Beaverton, OR 97006, USA
jay.melican@intel.com

Abstract. Drawing upon James Carey's ritual model of communication as a framework, we argue that rituals, especially religious rituals, are important resources for technology design. We suggest that a ritual view of ICT interaction represents an alternative and significant model for ICT development and evaluation, and that the observance of religious rituals affords researchers the opportunity to see cultural values at the peak of their expression. To illustrate, we describe several examples and three case studies of religious rituals that involve technology. For each, we discuss the ritual's enactment, where and how it intersects with technology, and the broader cultural values it embodies. We conclude with remarks about how religious values are meaningful for the design of culturally relevant consumer technologies and we offer advice on how other researchers can use ritual observation to inform and inspire their technology designs.

Keywords: Religion, ritual, user research, ethnography, information and communication technologies (ICTs), technology design, technology evaluation.

1 Introduction

In Kyoto, Japan, a high-tech executive makes a pilgrimage to the Buddhist shrine most appropriate to his current circumstances. He has just turned 41 and therefore he has begun his most inauspicious yakudoshi, *or unlucky year. With his camera-phone, he captures an image of the shrine he has come to visit. Days later, back in his hometown of Mito, he shares the image with his dinner guests.*

Anna Poggi, Milanese mother, homemaker and part-time, public-sector secretary, shows us with pride the contents of her subsidized apartment -- her treasured possessions include the doll she received for her First Communion, her wedding dress, and the cassette player that she was given on the occasion of her confirmation in the Roman Catholic Church.

At the southernmost edge of Taipei, dozens of Taiwanese families have gathered at the public mortuary facility to take part in one of the many Buddhist, Taoist and Christian funerals being held simultaneously within the complex. Family members

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 74–82, 2007.
© Springer-Verlag Berlin Heidelberg 2007

closest to the departed have covered their heads with white veils. They toss paper money, paper clothes and fashion accessories...even paper mobile phones into a large ceremonial incinerator.

As businesses struggle to compete in a global marketplace, they increasingly turn to ethnographers to inform and give direction to product design by shedding light on the most salient cultural values and concerns of unfamiliar consumers. Perhaps because of a perceived ill-fit or seeming irrelevance, technology researchers -- with a few notable exceptions [1-5], have failed to study religious rituals in the course of their fieldwork. We believe that this is a mistake; that by neglecting spirituality and ritual in technology research, researchers are missing the opportunity to study cultural values at the peak of their expression. Furthermore, we believe that researchers often overlook important functions of ICTs in the reification and maintenance of cultural and social values. To that end, we believe that religious rituals are a valuable opportunity for HCI researchers and designers to understand cultural values and to identify opportunities for technology design and innovation.

As demonstrated in the brief examples above, communication technologies and technological devices may play important roles where spiritual beliefs and values are put into practice. These interactions represent an important, and often overlooked, role that communication technologies are playing in the construction and maintenance of social life. Traditional (U.S. American) academic conceptions of communication and, by extension, interaction with communication technologies center on the transmission of information over time and/or distance for the purpose of inducing some kind of measurable change. As a result, ICT designs are often evaluated based on the extent to which they can produce a desired change – whether or not they let people communicate faster, with more precision, over greater distances, and so forth.

In this paper, we offer an additional perspective to understanding the role of ICTs in social life, based in a ritual view of communication [6]. Rather than focusing on the transmission value of ICTs, we consider ICT interactions primarily as they function to represent shared beliefs and maintain social order over time. We demonstrate this perspective in action with several examples drawn from our ethnographic field research. Specifically, we present three case studies, each involving ICTs in religious rituals. We use these examples to illustrate both how ICTs can be understood for their ritual value, and also to demonstrate how studying ICTs used in the course of religious ritual can offer critical insight into the core values and attitudes of a particular social and/or cultural group. We conclude with comments on methodological implications and advice on how other HCI researchers may use a ritual framework to inform their own technology designs.

2 A Ritual Model of Communication

In his 1988 book, James Carey argues that a ritual model of communication must complement more traditional transmission models if we are to fully understand the range of human communication acts. Although a transmission view of communication whereby signals or messages are sent over distance for the purpose of inducing change, was (and still is) the most common approach to understanding communication in the American scholarly tradition, it represents only a subset of the communication acts we

undertake every day. Many, if not most, of our communicative acts, Carey argues, are not directed toward the extension of messages across time and/or space, but rather "toward the maintenance of society in time," [6]. In other words, much of our communication is directed not at the creation of personal or social change, but rather at the maintenance of social and cultural order. It is therefore important to consider communication technologies not only as tools of transmission – but also for the part they play in the reification of the society and culture in which they exist.

Carey further argues that rituals serve as a "projection of community ideals," where cultural norms are embodied in an "artificial but nonetheless real symbolic order that operates to provide not information but confirmation, not to alter attitudes or change minds but to represent an underlying order of things," [6]. As rituals are enacted, social norms and cultural values become embodied in form and easily recognizable. The movements of a dancer, the thematic elements of a text, and even the digital images on a mobile phone all become symbolic representations of cultural values that are readily accessible for the HCI researcher to identify, observe, and understand.

A goal for technology researchers, then, can be to identify rituals involving ICTs for in situ observation and analysis. Carey argues that the vast majority of communication acts are ritual. Even instances of communication that may, at first glance, appear to be completely informative are often deeply ritualistic. For instance, newspapers may be considered a medium for disseminating news and knowledge, forming or changing attitudes, and identifying issues of public import. However, newspapers may also be understood within a ritual view of communication where:

> *"Reading a newspaper [is viewed] less as sending or gaining information and more as attending a [religious] mass, a situation in which nothing new is learned but in which a particular view of the world is portrayed and confirmed. … We recognize, as with religious rituals, that news changes little and yet is intrinsically satisfying; it performs few functions yet is habitually consumed. Newspapers do not operate as a source of effects or functions but as dramatically satisfying, which is not to say pleasing, presentations of what the world at root is," [6]*

However, when pressed for time, as we often are when we are on field research assignments, the rituals woven into daily life and routine social interaction may be difficult to quickly identify and meaningfully analyze for their potential to inform ICT development. It is for that reason that we argue for the study of religious rituals involving ICTs as a rich resource for technology research and design. As Carey suggests:

> *"If the archetypal case of communication under a transmission view is the extension of messages across geography for the purpose of control, the archetypal case under a ritual view is the sacred ceremony that draws persons together in fellowship and commonality." [6]*

By examining the most archetypal examples of communication rituals – religious rituals - and the ways that ICTs become involved, we are offered a glimpse into the intersection of technology and cultural values at the height of their expression. These religious rituals – weddings, funerals, birth rites, and so forth – are often easily identifiable, and in

our experience, quite accessible for study. Furthermore, the customs and symbols associated with these rituals are often consistent over time, widely recognized within the population, and easily articulated. As a result, it is not only easy to access religious rituals for study, but also it is easy to find people who can talk about religious rituals in meaningful ways. Furthermore, these rituals increasingly involve ICTs in some way – often to document, support, mediate, or otherwise inform the ritual as it is enacted. So, because religious rituals embody cultural values, are easy to study, and frequently involve ICTs, they can be used by technology researchers in two ways: to quickly identify important social and cultural values as they relate to ICTs, and to identify opportunities for designing ICTs based not only on their ability to transmit information but also their ability to enact, document or otherwise support ritual.

3 Methods

The case studies that follow are drawn from data collected in the course of a 6-month, 12-country, exploratory research project designed to explore the social and cultural construction of personal and domestic technologies, including mobile phones, personal computers, and televisions. Teams of 2-3 researchers spent a minimum of 2 weeks visiting and interviewing families in each of the 12 countries studied. In each country, the researchers worked with a resident interpreter to identify 7-12 primary participants for a variety of research activities. Interpreters translated all research interviews not conducted in English. At minimum, each participant was interviewed at home a single time. Most participants were visited multiple times at home and also shadowed at their places of work, at social events, and at family gatherings. Additional participants were recruited to complete participatory design activities, cultural probes, photo diaries, and mapping exercises. In total, over 200 participants completed one or more research activities in the course of this study.

The data described below were collected in Taipei, Taiwan, Mito, Japan, and Milan, Italy. In all cases, participants were aware that they were being observed or interviewed by Americans working for Intel Corporation. The Italian participant was directly recruited for an interview and was compensated for her participation. The Japanese participant was interviewed during the course of a social event, and was not compensated. The Taiwanese participants were observed but not interviewed, and they were not compensated. All participant names have been changed to protect their privacy.

4 Cases from Ethnographic Research

We introduced this paper with three short vignettes that offer an alternative perspective based in a ritual view of communication and consider ICT interactions primarily as they function to represent shared beliefs and maintain social order. Below, we elaborate on those and other examples drawn from our ethnographic field research, and we build on Carey's model of ritual communication by proposing a theoretical framework for interpreting the many dimensions in which ICT interactions can be understood and modeled as rituals of communion and communication. First, we

briefly consider technology interactions that are central to ritual enactment - those that support ritual communication or the performance of a ritual itself. Next, we consider interactions in which technology's primary function is reproductive, in which ICTs are used to document the performance of ritual and/or memorialize the state of the actor as s/he embarks upon or completes a rite of passage. Finally, we consider interactions in which technological artifacts themselves become elements in the symbolic economy of the ritual.

4.1 ICT Interactions as Supporting and Extending Ritual

Perhaps the most obvious and most well researched cases of religious uses of modern information and communication technology are examples in which technological interactions are central to communal ritual enactment, functioning as mechanisms of mass communication: the addition of loud speakers to mosque minarets to amplify the call to prayer [1], televised church services, and modern mega-churches where amplification, display and broadcast technologies permeate the worship service itself and extend communication with the congregation well beyond scheduled service hours [4]. In these cases, technology has functionally supplemented or even supplanted the elaborate awe-inspiring construction of gothic cathedrals and great mosques and temples; the screens and loud-speakers that dominate mega-churches not only play a central role in communication to ever wider audiences across time and space, they also stand in as signs of a religious community's affluence, expansiveness and potential influence.

ICTs may also be central to religious observance outside of communal services, as, for example, when computing tools facilitate control of the domestic environment and enable performance of ritual duties by proxy. One example can be found in Woodruff, Augustin, and Foucault's recent ethnographies of Modern Orthodox Jews living in Brooklyn, New York. The PC-controlled, X-10 home automation systems that control their lights and electrical home appliances allow these modern families additional comfort and piece of mind in their recognition of the Sabbath as a day of rest and reflection and their strict observance of *Halacha* – the body of rules that forbids, among other things, the manipulation of electrical circuits on Saturdays [7]. Another manifestation of such ICT support can be found in long-distance performance of rituals in diasporic religious communities. It is not uncommon that Turkish Muslims, for instance, may migrate – temporarily or permanently -- from their region of origin in a quest for stable employment. Internet-based services now permit displaced Turks to pay respects to deceased relatives by hiring a camcorder-toting representative to visit family graves in their stead. Through another online service (provided by larger grocery store chains), they may purchase a lamb for sacrifice on the holiday of *Kurban Bayrami* (*Eid el-Adha), and* distribute its meat amongst relatives and charitable organizations.

Such ICT interactions – in which technologies support the performance of rituals or extend the distance from which rituals of communion can be preformed – come closest to conforming with tradition transmission models of mediated communication. They are perhaps most noteworthy as indicators of the acceptance and absorption of technologies of mediation in what many might consider sacred realms of activity.

4.2 ICTS Interactions and the Memorialization of Ritual

The story of Futo-san, a high-tech executive living in Mito (outside Tokyo), Japan, provides an example of yet another common – though less commonly studied – aspect of ICT use in religious ritual. We met Futo-san informally over dinner with colleagues in Mito. After the obligatory exchange of business cards, the conversation quickly shifted to the topic of our research – Japanese social conceptions of mobile phones. Futo-san explained that he was most grateful for the camera on his mobile phone as it allowed him to document personally meaningful experiences. When we probed for more information, he showed us the photo above in Figure 2. He explained that this photograph is a personal reminder of his recent pilgrimage to Kyoto, taken to ward off the ill-effects of his *yakudoshi* or unlucky year. Although the photograph is normally private and rarely looked at even by Futo-san himself, it is nonetheless an important reminder of his journey. Furthermore, because the photograph is on his mobile phone, it is portable and always close at hand, extending the benefits of his pilgrimage to protect him throughout the year.



Fig. 1. Futo-san sharing his story with Intel researcher Todd Harple



Fig. 2. Image of the Kyoto temple from Futo-san's mobile phone

Futo-san's story describes the marking of a significant life-stage transition, the passage -- across a boundary line deeply inscribed by religious tradition -- from one territory of spiritual experience to the next. In making his ritual pilgrimage, Futo-san embodied this spiritual passage in his physical movement (to Kyoto) and his drive to find his life's path through reflective discipline. His use of a camera phone to memorialize the journey embodies an important Japanese cultural value – privacy and modesty. While most photographic portraits of individuals on occasions of religious import (weddings, baptisms, etc.) are taken and reproduced primarily for purposes display – to communicate to family or to visitors to a family's home, Futo-san's photo captured and stored on his camera-phone was intended primarily for his own reference. It is entirely personal, never intended to transmit information to anyone else, yet it is deeply meaningful and completely appropriate within Futo-san's culture and society.

4.3 ICTS as Signs in Ritual

The final two examples included in the introduction to this paper represent variations on a single theme and serve as illustrations of ICT interactions in which the technological artifacts themselves function as signs within the symbolic economy of religious rituals. Anna Poggi, the 43-year-old, Milanese mother of two described above does not consider herself to be a particularly devout Catholic (no more so than her friends and neighbors). Yet, in her home she displays her First Communion portrait and the doll she was given as a gift on that occasion long ago. She has also kept and has now given to her four-year-old daughter a long-outmoded audio cassette player



Fig. 3. Anna shows a photo portrait and doll from her First Communion



Fig. 4. The audio cassette player Anna received at her Confirmation

that she received as a gift for her Confirmation in the Catholic Church. Though the artifact has no functional value to her (and little to her daughter, who has two, more modern cassette players), it has symbolic value and is thus worthy not only of retention beyond its logical usefulness but also of gifting. Just as a wrist watch gifted to a young Muslim boy at his circumcision or high school graduation might be worn through adulthood and passed on to children and grandchildren, Anna's "boom-box"-style cassette player has taken on value beyond its ability to play music. Just as the photo portrait memorializes her First Communion, the boom-box represents (mostly to her and to the family members who gave it to her) the ritual itself – Anna's confirmation as an adult member of her religious community. Significantly, these artifacts on display in her home (including her cassette player) cue Anna's recounting – to visitors and, presumably to her children -- of her family's history, her religious traditions, and her personal experiences with her church and progress through it's various rituals and rites of passage.

Perhaps the most striking example from our studies of technology with great ritual value but no functional transmission value is the example of families burning "ghost" or "spirit" mobile phones and laptops for deceased relatives at Taiwanese funerals. The photos above depict a scene common to regions of the world with Buddhist religious traditions. Friends and family members of the deceased burn "spirit money" and other necessities, here for recently dead, but on other occasions for long-dead loved ones.



Fig. 5. Spirit money, paper ICTs and assorted afterlife accessories



Fig. 6. Family members of the deceased burn offerings at a crematorium in Taipei

The paper mobile phones, PDAs, televisions, laptops, and battery recharging stations that are now offered alongside paper money, mahjong sets, make-up kits and clothing are testament to the symbolic value of the signified, real-world technological artifacts. Even those ancestors who never used a mobile phone or laptop when they were living are afforded these luxuries in the afterlife through religious ritual. As one Taiwanese woman explained to us, to deny your ancestors the things that you enjoy would be disrespectful. It is the obligation of the living to ensure the deceased enjoy all of the luxuries f modern life.

5 Discussion

As the examples above illustrate, ICTs are frequently, and meaningfully, implicated in the enactment of religious rituals. With each example, we have described how deeply-held cultural values are exposed in the course of the ritual performance, and how the relationships between technology and ritual offers important clues about the social and cultural meaning of each technology involved. We hope these examples have sufficiently convinced fellow technology researchers to consider religious rituals in the course of their future fieldwork. However, the goal of this paper was not only a methodological intervention. We also argue for the adoption of a ritual view of ICTs to supplement the more traditional transmission view of ICTs. Plainly, designing and evaluating ICTs for their abilities to transmit information across time and space are credible and important goals. However, transmission represents only a portion of human communication function and need. Ritual – religious or otherwise – is an equally important communication goal and therefore ought to be acknowledged and considered in our ICT design and evaluation frameworks as such.

In the course of our fieldwork, we encountered dozens of examples of technology being used in the course of religious rituals. We selected the three described here because they clearly represent the failings of a transmission-only model for understanding the role and social function of ICTs. In each case, from a transmission-only perspective the ICT is non-functional, minimally-functional, or functional in a way other than that intended by the designer. But, in each case, the ICT is nonetheless playing a meaningful role in social life – a role that is desirable, value-centered, and widely understood. As researchers and designers sent to the field to collect data to inform ICT design, it would be easy to ignore instances where technology seems to be

operating outside of its technically-prescribed role. However, if we did so, we would overlook opportunities to design for a full range of communication needs.

As we have demonstrated, photographs can be un-shared, yet deeply meaningful; music players can be outdated, yet still cherished; and mobile phones can be non-functional, yet still desirable. Moreover, we believe that, equipped with this framework for understanding the ritual value of ICTs, future developers can create technologies that artfully combine both communication models – photographs that remain private, but can be shared when the time is right; music players that are up-to-date, but remain nostalgic; and mobile phones that are fully functional, yet still appropriately respectful and reverent. To that end, we believe this framework has value beyond the domain of religious ritual. It extends to apply to many types of rituals – including those enacted in the course of everyday life. As Carey suggested in his example of the ritual value of newspaper reading, many, if not most, human communication activities can be understood both for their ability to transform culture and their ability to maintain it. We believe that by considering the ritual value of technology, many instances of "idiosyncratic" or "incorrect" technology use are explained. Suddenly, within a ritual framework, behaviors that baffle technology researchers – carrying a mobile phone that does not have a service plan, buying a computer that never gets turned on, taking digital photos that are never looked at – make perfect sense. Although information is not transmitted in any of these cases, rituals are enacted. Our hope is that using the ritual perspective introduced in this paper technology researchers can not only make sense of human orientations to ICTs, but also make recommendations for ICT design and evaluation that acknowledge and support the full range of desired and desirable human communication behaviors.

Acknowledgments. The authors would like to thank their Intel colleagues, most especially Genevieve Bell, Sue Faulkner, Todd Harple, and Allison Woodruff for their help with the examples contained within this paper and with earlier versions of this analysis. We would also like to thank Justine Cassell and our anonymous reviewers for their helpful comments. Finally, we extend our gratitude to the interpreters who helped us conduct this research, and to the participants who informed this work.

References

1. Bell, G.: The Age of Auspicious Computing? Interactions 11(5), 76–77 (2004)
2. Hoover, S.M., Schofield Clark, L.: Practicing Religion in the Age of the Media. Columbia University Press, New York (2002)
3. Muller, M.J., et al.: Spiritual Life and Information Technology. Communications of the ACM 45(2), 82–83 (2001)
4. Wyche, S., et al.: Technology in Spiritual Formation: An Exploratory Study of Computer Mediated Religious Communications. In: CSCW 2006 Banff, Canada (2006)
5. Zaleski, J.: The Soul of Cyber Space: How Technology is Changing Our Spiritual Lives, San Francisco, Harper Edge (1997)
6. Carey, J.W.: A Cultural Approach to Communication. In: Carey, J.W. (ed.) Communication as Culture: Essays on Media and Society, pp. 13–36. Routledge, New York (1988)
7. Woodruff, A., Augustin, S., Foucault, B.: Sabbath Day Home Automation: "It's Like Mixing Technology and Religion". In: Proceedings of the 2007 Conference on Human Factors in Computing Systems (CHI2007), ACM Press, San Jose, California (2007)

Shanghaied in a User-Friendly Manner - An American's Initial Experiences in a Full-Time Usability Job in China

Brian I. Glucroft

HFI China, 407, No. 555 West Nanjing Road,
200041 Shanghai, People's Republic of China
brian.glucroft@humanfactors.com

Abstract. As the application of user-centered design spreads across the globe, technology companies are facing new challenges in establishing usability teams in non-western countries. Managers must decide whether to staff their usability teams with local or foreign individuals, and this decision can be influenced by the availability of usability experts who are native to the country. China's rapid economic growth has led to a strong demand for usability practitioners. Given the relatively small size of the usability community in China, there are unique opportunities for non-Chinese nationals. In this paper, I describe the initial experiences I faced as an American joining a usability team of Chinese nationals. I discuss my preparation and experience before arriving in China, as well as the adjustments I had had to make while conducting user-centered design in a culture that was very different from my own. I believe the sharing of my experiences in both work and non-work settings can offer helpful insights to other non-Chinese nationals interested in conducting usability work in China, as well as to managers who are considering adding non-local staff to their usability team.

Keywords: working abroad, user-centered design, China, cultural adaptation.

1 Introduction

The practices of user-centered design have been steadily spreading across the globe. With its vast population, rapid economic growth, and its recent embrace of various technologies, China in particular is a prime location for the rapid infusion of new usability practices. However, a usability community does not grow overnight. A mature presence requires a mix of industry expertise, academic research, and higher level training. While China has been fostering a small usability community for a number of years, its current needs may be outstripping the current availability of experienced professionals. For that reason, there are rich opportunities for non-Chinese nationals to make important contributions alongside their Chinese colleagues.

After having studied Chinese language for 2 years in the US, I accepted a position with HFI China in Shanghai. I was attracted to the job because I welcomed the additional challenge of experiencing life in a very different culture. My brief time so far in China has been far more than I ever could have imagined.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 83–88, 2007.
© Springer-Verlag Berlin Heidelberg 2007

I believe that in sharing some of my experience, others considering a similar career move and those involved with hiring people can gain a glimpse into some of the issues faced by a Westerner coming to work in China. I realize my experiences may not be representative or complete, however, they offer a chance to see the flavor of such an experience and an insight into the some of the contributions that can be made under such circumstances.

Many people around the world work outside of the country where they have spent the majority of their life. However, for those in the usability domain, working outside of one's native country poses a unique challenge given the importance of understanding the user. Can an outsider really gain the same level of understanding of a user as a local person? Are there in fact any advantages of being non-local? These are all questions I hope to explore as I continue my work in China. However, already I feel that there is something to share.

2 Before China

Preparations for moving abroad can be very daunting. Everything from deciding what and how to move to making sure that one leaves things in a proper state back "home" can add up to a lot of time and stress.

Since I was new to HFI and their office in China was new, I was asked to spend 6 weeks in India familiarizing myself with HFI as I participated in their training course for usability professionals. I embraced the opportunity to spend time in a new country and made sure to arrive two weeks early in order to travel around the country and acclimate myself.

While the focus of the trip was primarily to familiarize myself with HFI's practices and people, I found that the incredible India experience was what taught me the most. Working in India for even that short period of time required me to adjust quickly. India was an interesting paradox: I had few issues with the language, given the prevalent use of English, and yet the culture was strikingly different from anything I had ever known.

Whether it was understanding the commonly used head waggle that I had never seen before, or learning how to explain to someone that something had to be done (and done soon), my experience in India served as a sort of practice grounds for what I would later experience in China.

Another benefit of the trip to India is that there is currently a lot of interaction between HFI's India offices and HFI China. While none of the work I have done so far has directly pertained to India, my experience there has allowed me to better understanding and appreciation of my Indian colleagues. While of course 8 weeks in India does not make one a master of Indian culture, my current understanding is worlds apart from what it would have been otherwise.

3 Work in China

After my time in India I arrived in Shanghai. Numerous mundane challenges awaited me like obtaining necessary permits/visas, health exams, and finding a place to live.

Now, I could finally immerse myself into the experience of living and working and China.

While I had ideas, I was not sure in what ways I would be able to fit in to a team comprised entirely of Chinese while working on Chinese projects. Like any new job, there was a period of acclimatization that was only more pronounced due to the large change in environment. However, I quickly jumped head first into projects and began to discover how I could make a unique contribution to the team.

3.1 Disclosure of a One-Way Mirror

One example involved the testing practices for conducting a focus group. The focus group occurred in a room which was viewable through a one-way mirror. In previous practice, participants were told they were being observed but were not explicitly told there was a one way mirror. Upon understanding the procedure, I insisted that participants should be informed of the one way mirror.

My colleague insisted that if the participants were informed of the one way mirror they would become uncomfortable and unable to focus on the interview. She explained that since we were in China we had to adjust our methods.

I explained that I felt it would be unethical not to inform the participants of the one-way mirror. If this affected our data then we simply had to find another solution or live with it. My colleague still protested.

I then explained that although US participants may be more used to such a situation, there were still many individuals who would fit the stereotypes of Chinese participants in terms of their degree of frankness or comfort in such situations. In the US we didn't simply throw them out or change procedures, but we would try to find ways to work with such individuals if possible.

I questioned whether, from a participant's view, the ethical issues involved might be different in China. I asked my colleague to imagine herself as a focus group participant and that she had not been informed that there was a one-way mirror. I then asked her to imagine that after the testing she was informed about the one-way mirror. How would she feel? My colleague came to a quick conclusion that this would upset her and she would not feel good about the situation. I replied that her response seemed to support the notion that Chinese participants would expect to be informed of such issues. She agreed but continued to voice concerns that it would still lead to the participants being uncomfortable during testing.

I continued to insist that we simply couldn't give up and sacrifice our obligations as researchers. It might not be simple, but it was our responsibility to find a solution that provided the best research conditions for us without sacrificing the rights of the participant.

After about 5 minutes of thought, my colleague said she had an idea. She suggested that she introduce me before the testing, I make a few comments in Chinese, and she then point out the one-way mirror and explain that I would be observing. She felt this might create an environment where we could inform the participants of the one-way mirror without making them uncomfortable. I replied that I would be happy to do anything that would enable us to properly inform them of how they were being observed.

We carried out the plan as described. When I spoke a little Chinese to the participants they responded extremely warmly and clearly appreciate my gesture. The group seemed very relaxed and had no questions about the observation process. The testing went smoothly. There was never any indication the group was uncomfortable and in fact it was extremely forthcoming in the discussion.

Afterwards, my colleague and I agreed it was a success.

3.2 The Dialects of China

In another instance, I faced the issue of dealing with a specific dialect of China. Although Mandarin is the national language of China, Mandarin, many parts of China retain the use of local dialects. These dialects are often, from a practical standpoint, as different from Mandarin as another language.

For a project in Wuhan, China, we initially had a simultaneous translator from Wuhan. Given that our moderator did not speak Wuhanese, all participants were asked to speak in Mandarin. After a period of time it became clear that the translator was not sufficiently skilled for our needs so a new translator was quickly flown in from Shanghai. While the new translator was clearly superior, we ran into some problems when participants sometimes slipped into speaking Wuhanese. The previous translator had been able to effortlessly translate portions that were spoken in Wuhanese while the "better" translator could not. This did not have a large effect on the actual interview process since the moderator was restricted to Mandarin no matter what. However, for later data analysis it proved useful to have translations for the bits that were spoken in Wuhanese.

This experience made me wonder whether we could have gotten more expressive or detailed replies to questions had the interview been conducted in the local dialect. The lesson for me was that while Mandarin is the national spoken language, it does not mean it will be participants' language of choice. Depending on the nature of the participants and the requirements of the research, using the local dialect could be advantageous or critical.

3.3 Translating Culture

Sometimes, there is more to be translated than spoken or written language. The implications of specific body language, tone of voice, mannerisms, etc. can differ across cultures. The next case presents such an example.

An Indian colleague and I were observing a focus group from a separate room. At the beginning of the session a man sitting close to the moderator suddenly stood up. In a loud voice he began to complain that the session was being conducted in Mandarin and not in the local dialect. It was not long before the Indian colleague stood up and said she was going to enter the room with the focus group to resolve the situation. I immediately recognized how the situation appeared to her and quickly explained that it wasn't as bad as it seemed and that we needed to let the moderator handle the situation. The moderator proceeded to calmly speak to the man and explain that since she was not from the local area she had to conduct the interview in Mandarin. After a minute or so the man sat down and the session went on without a hitch.

The man's reaction would have been viewed as extremely threatening in US (and apparently Indian) culture. Had the testing occurred in the US I would have quickly entered the room to assist the moderator. However, in China it is not uncommon for people to express themselves in such a manner without it having the same overtones it would have in the US. My experience not only led me to be able to quickly place the man's reaction in a proper context, but to also quickly recognize how the other observer was viewing the situation. Having a clearer picture of both sides enabled me to effectively head off an unnecessarily strong intervention. It was a case where not only my understanding of Chinese culture, but also my understanding how Chinese culture might be perceived by others played a key role in allowing a process to run as efficiently as possible.

3.4 Applying Personal Experiences to Work

Living in China gives me the chance to explore and take in the culture at a much deeper level than I could by simply traveling here for a 1 week project and departing. Given my natural tendency to conduct "amateur ethnography" when I travel about, I quickly accumulated a wealth of knowledge and intuitions about Chinese culture. Working with Chinese colleagues and interacting with Chinese friends was not only an additional experience but also an opportunity to see how closely my view of things matched theirs.

These experiences proved invaluable on a project examining owners of small groceries and convenience stores in another Chinese province. My previous exposure to such stores in China gave me an advantage compared to simply coming straight into the country from scratch. However, my familiarity with stores outside of Shanghai (and China) kept me from falling into the trap of assuming practices in the other province would match those of Shanghai (in fact, in many ways they did not). Having a sense of bargaining habits, products available in stores, how money was transacted, etc. allowed me to better situate myself into the domain of the project.

The point here is simply that for usability professionals in a foreign country, daily non-work experiences which are only available to those living in the country can play a significant role in facilitating their work.

3.5 Office Culture

Prior to arriving in China I had read several books on Chinese business culture. They were very intriguing and I look forward to applying what I had learned in my everyday actions with colleagues.

However, when I arrived I quickly realized that while the books may represent a certain segment of people, they did not always apply to the people I interacted. Many of my colleagues were under 30. This age group is very different from older generations. Their lives have been shaped by events that differed significantly from those of their parents. They have grown up in a more "westernized" China. I found it much easier to work with them than the books led me to expect. On a regular basis I would find their actions directly contradicting what a book would have predicted. The books were still valuable as providing a reference point from which to interpret actions, but the books were simply not directly applicable to my situations.

Even seemingly simple issues such as defining the number of days off for a holiday can require a great deal of effort to sort out. In China, for a holiday it is common for companies to shift some of the workdays to a preceding and/or following weekend in order to provide a bigger block of consecutives days during which the holiday is closed. However, I found I had to be extremely explicit in what I meant when I asked how many days I had off. If I asked the question (in English) the answer often corresponded to the number of consecutive days the office was closed. However, what I was really interested in was the number corresponding to the reduced number of days I would actually have to work. The misunderstanding wasn't a simple language problem, but that the typical Chinese model and American model of holiday structure were so different from one another that both sides didn't appreciate how the other side was interpreting and framing comments.

Much is often made about the differences in ability to provide criticism in Chinese and American culture. While certainly this is true, especially when coupled with issues of hierarchy, I found that for project work I had little difficulty with these differences.

4 Conclusion

My experience in China has only just begun. However, I have already seen how a foreigner can make significant contributions in applying user-centered design in China. As my experience grows, I hope to find more ways to combine my set of perspectives and skills with those of other professionals. Ultimately, I believe that a mix of "insiders" and "outsiders" can lead to the most effective usability solutions.

Every day when I am walking around China, I feel as if I am doing ethnography. This outlook gives me an extra boost of energy, helps enable me to overcome the challenges of working in a "foreign" culture, and provides me with a rich set of additional information to apply to my work. Of course, reading books can assist in beginning to understand another culture, but especially in a fast changing place like China, living there can provide a much deeper level of understanding. Although foreigner's knowledge of the local culture is unlikely to reach the same point as a local person's, he or she is also able to offer a new perspective. Complementing a local awareness with a foreign awareness can bring a different set of ideas and perspectives to the table while developing an appreciation of the culture under study.

Acknowledgments. I would like to thank Apala Lahiri Chavan, Eric Schaffer, and Kath Straub for their efforts in bringing me to both India and China and making this "experiment" possible. I would also like to thank Uyen Le for her review of this paper.

A Tool for Cross-Cultural Human Computer Interaction Analysis

Rüdiger Heimgärtner

Siemens AG, Im Gewerbepark C25, 93055 Regensburg, Germany
Ruediger.Heimgaertner@siemens.com

Abstract. This paper describes a tool for analyzing cross-cultural human computer interaction (HCI). From literature and reasoning possible cultural HCI indicators have been identified and measured with this tool to compare them in respect to the different culture of the users. Concept, implementation, usage, benefit and implications of this tool will be presented. Two online studies using this tool concerning cultural adaptability exemplified by use cases of navigation systems revealed differences in interaction behavior that depend on the cultural background of the users (e.g. attitude, preference, skill etc.) and proved that the tool is working properly.

Keywords: cultural adaptability, cultural user interface design, adaptive HCI (Human Computer Interaction/Interface), HMI (Human Machine Interaction/ Interface), cross-cultural HCI analysis, driver navigation systems, tool.

1 Introduction

The "Intercultural Interaction Analysis" tool (IIA tool) was developed to obtain data regarding cultural differences in HCI simulating use cases – in this case, navigation systems [6]. The main objective of the IIA tool is to observe and analyze the interaction behavior of users from different cultures with a computer system to determine different interaction patterns according to the cultural background of the users. Culture influences the interaction of the user with the computer because of the movement of the user in a cultural surrounding [13]. To locate and find out the kind of different interaction behavior of the users from different cultural groups (at national level (country) first because of the high cultural distance) the interaction behavior of the users with the computer will be observed and detected. The objective is to be able to draw inferences *regarding differences of the cultural imprint* of users *by analyzing the interaction behavior* of those users with a computer system to get knowledge that is relevant for intercultural user interface design and a necessary precondition for cultural adaptive systems [7]. E.g. the right number and arrangement of information units is very important for an application whose display is very small and at the same time the mental workload of the user has to be as low as possible (e.g. driver navigation systems).

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 89–98, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 A Tool for the Analysis of Cultural Differences in HCI

2.1 Basic Reflections

Measuring the User Behavior. Most methods to measure the user behavior are located in the domain of psychological research. They are designed with respect to intercultural phenomena concerning human-human-interaction and intercultural communication. Measuring user behavior in cross-cultural HCI has not been investigated in such detail. Therefore, it is necessary to perform research in this area by introducing new methods such as analyzing critical interaction situations between humans and computer or machines, computer-supported cooperative work or cognitive technology research [6]. One of the most promising approaches is to employ usability metrics, because of its empirical value [15] [2]. The objective is to find cultural differences in quantitative metrics for HCI by relying on existing metrics in HCI and developing new metrics that are adequate to measure cross-cultural HCI. Thus, the IIA tool has been designed and used to conduct two studies to get new findings in this area.

Use Cases of Cross-Cultural HMI Design in Automotive Navigation Systems. In order to limit the scope of research, representative and demonstrative use cases have been restricted for cross-cultural HMI in automotive navigation systems. The most interesting use cases possess a high degree of interactionality. One such significant use case is voice guidance: the driver receives subsequent information as to when and in what distance a driving maneuver needs to be performed to reach a predefined destination. In Japan, it is common in current automotive navigation systems to present this piece of information to the driver very frequently. In Germany, this would be felt as information overload. This difference can be explained according to Hall [3], who claims that every culture has its own speed of communication as well as to Hofstede's cultural dimension of uncertainty avoidance [9]: Japanese are worrying more about not having enough information than Germans [12]. Relationship-oriented cultures accept high information speed in contrast to task-oriented cultures that like to concentrate on their tasks instead of wasting time with communicating information, which is not task-oriented such as chatting with other people [5] [3]. Another use case is destination input. Which input methods does the user prefer? In China, input method editors have to be used because of the numerous Chinese characters. Yet another use case is map display. What map direction is best according to the user's cognitive style? How many points of interest (POI) should be presented to the user? Task-oriented users prefer fulfilling tasks to relationships (during working hours, e.g. professional drivers). The knowledge, whether the user is more relationship-oriented or more task-oriented, may be derived from the user interaction behavior. For example, pressing buttons very exactly and navigating very directly, without permitting disturbances or interruptions by other people or the system, increases the probability that the user is task-oriented because he takes the task very seriously.

Implementation of Test Tasks. Using the method of literature research and analytical reasoning, 118 potentially culturally sensitive parameters have been identified, implemented into the IIA tool and applied by measuring the interaction

behavior of the test persons in relation to the culture [6] [7]. The IIA tool consists of three elements: a data collection module, a data analysis module and a data evaluation module for cross-validation of data. To test some of the most important interrelationships between culture and information processing, the IIA tool allows the measurement of numerical values like information speed, information density, information context and interaction speed in relation to the user. These are hypothetically correlated to cultural variables concerning the surface level like number or position of pictures in the layout or affecting the interaction level such as frequency of voice guidance. Thus, a hypothesis like "there is a high correlation of high information density to relationship-oriented cultures such as China" should be confirmable by adjusting more POI by Chinese users compared to German users. So, the use case "map display" was simulated by the map display test task to measure the number of pieces of information on the map display regarding information density (e.g. restaurants, streets, points of interest (POI)) (Figure 1).

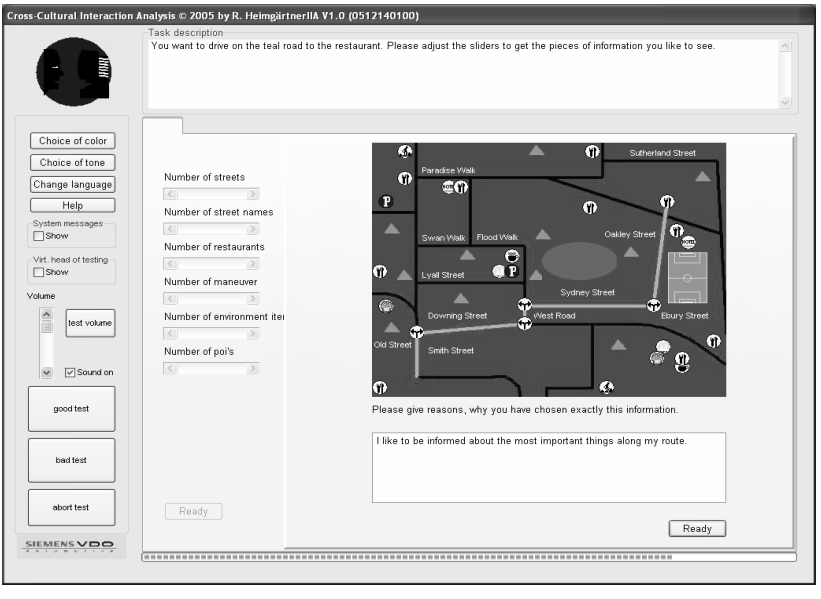


Fig. 1. Map display test task

Every one of the test tasks serves to investigate other cultural aspects of HCI. E.g. the special use case "maneuver guidance" has been implemented into the test task "maneuver guidance" where the test user has to adjust the number and the distance of the maneuver advice messages on the screen concerning the frequency and speed of information. Both abstract and special use cases have been implemented in this way as test scenarios into the IIA data collection tool in order to obtain results for the design of navigation systems [6]. Based on this principle, this test tool can also be used to investigate the values of other cultural variables like widget positions, menu structure, layout structure, interaction speed, speed of information input and output, dialog structure, etc. The test tasks (use cases) are localized but designed semantically identical for all users: they can be done by users of many different cultures.

IIA-Tool Setup, Test Setting and Usage. A user test session with the IIA tool comprises five parts: collection of demographic data, test tasks, VSM94 questionnaire, evaluation of results by the user, and debriefing questionnaire. The demographic questionnaire part delivers knowledge about the cultural background of the user (mother tongue, languages, nationality, residence in foreign countries). The developed and implemented test tasks in the IIA tool serve to motivate the user to interact with the computer and to test previously postulated hypotheses. To analyze the cultural values of the users, the value survey module (VSM94) has to be filled in by the user [10]. The VSM94 contains 26 questions to determine the values of the cultural dimensions using the indices from Hofstede that characterize the cultural behavior of the users [9]. The results of the VSM94 and of the test tasks are presented to the user who then has to estimate whether or not the cultural and informational values found correlate or match to him. The debriefing part reveals the purpose of the test to the user in detail and collects data regarding the usability of the test system, the perceived difficulty of the test in general and if the implemented hypotheses in the test tasks have been recognized by the user during the test session. During the whole test session, the IIA tool records the interaction between user and system, e.g. mouse moves, clicks, interaction breaks, or the values and changing of slide bars set up by the users in order to analyze the interactional patterns of users of different cultures.

2.2 Study and Data Collection

A local heuristic pre-study with seven Chinese and eleven German students served to check the intercultural usability of the IIA tool. This qualitative offline study showed the first interesting results regarding cultural dependent differences in interacting with the IIA tool. Some of the results have been confirmed by two online studies that have been conducted subsequently to verify the functionality and reliability of the IIA tool: Employees from SiemensVDO were invited per email to download the IIA tool and to do the test session. Table 1 characterizes the two online studies regarding total sample size, total tests downloaded, tests aborted, valid test data sets and total return rate.

Table 1. Setup of the Online Studies with the IIA tool

Study	*Sample size*				*Survey period*	*Total tests down-loaded*	*Tests aborted*	*Valid test data sets*	*Total re-turn rate*
#	Total	China	USA / Canada	Ger-many		#	in %	#	in %
1	600	200	200	200	12/14/05 - 01/14/06	166	41,5	102	16,6
2	14500	1500	4500	8500	11/14/06 - 01/19/07	2803	66,8	916	6,3

The tests have been aborted due to the following reasons: download time too long, no time to do the test now, test is not interesting or appealing. This type of qualitative data can help to optimize the testing equipment or to steer the direction of data analysis simply by asking the user for the reasons of his behavior during the test (e.g. using open questions via text boxes). Only complete and valid data sets have been analyzed using the IIA data analysis tool and the statistic program SPSS. The discrimination rate of classifying the users to their selected test language by the variables concerning the cultural background of the user (mother tongue, nationality, country of birth and primary residence) is 83.3% for the first and 81.9% for the second study.[1] Therefore, the differences in HCI in this study have been analyzed in relation to the three groups of test persons according to the selected test languages (Chinese (C), German (G), and English (E)).

Study Results: Cultural Interaction Indicators. The pre-study indicated that there are differences in the interaction between C and G with the computer regarding the order of pictures (more ordered by G than by C), test duration (longer for C), error clicks (C more than G) and telling the truth regarding computer experience (C understated their experience pretty much). One-way ANOVA, which is a statistical method to compare the means of more than two independent samples, was used in the two online studies to get significant cultural differences in variables, which are distributed normally. The results of the test of homogeneity of variances indicate whether ($p>.05$) or not ($p\leq.05$) the variables are distributed normally. A third of the potential variables is distributed normally and hence analyzed by ANOVA. Some of the variables show significant differences, which therefore can be called *cultural interaction indicators*. They represent significant differences in user interaction due to the different cultural background of the users (Table 2).

Table 2. Significant Cultural Interaction Indicators

Cultural interaction indicator	*First study*	*Second study*
Speed (MG)	F(2,102)=8,857**	χ^2 (2,916)=29,090**
MessageDistance (MG)	F(2,102)=7,645**	F(2,916)=16,241**
POI (MD)	F(2,102)=3,143*	χ^2 (2,916)=32,170**
MaximalOpenTasks	χ^2 (2,102)=12,543**	F(2,916)=15,140**
MaximalOpenTasks ratio (C,G,E)	2.5 : 1.4 : 1	1.7 : 1.03 : 1
Information speed value	χ^2 (2,102)=17,354**	χ^2 (2,916)=82,944**
Number of chars	χ^2 (2,102)=16,452**	χ^2 (2,916)=67,637**

The interactional differences between the user groups separated by the test languages have been identified using the Tukey-HSD-Post-Hoc-Test after one-way

[1] The discrimination rate and the standardized coefficients of the canonical discriminance functions in the brackets have been calculated using discriminance analysis (cross validated and grouped, Wilk's Lamda in study 1: $\lambda_{1\text{-}2}$=.072**, λ_2=.568**, Wilk's Lamda in study 2: $\lambda_{1\text{-}2}$=.192**, λ_2=.513**). The level of significance is referenced with asterisks in this paper (* p<.05, ** p<.01).

ANOVA. For the remaining variables, which are not distributed normally, Kruskal-Wallis-test has been applied. The variables in the valid test data sets are not distributed comparably in the first and second online data collection. Therefore, the same variables have been analyzed either by ANOVA or by Kruskal-Wallis-test (indicated with F or χ^2 in Table 2): *Speed (MG)* means the driving speed of the simulated car in the maneuver guidance test task ((C) less than (G) and (E)). *MessageDistance (MG)* denotes the temporal distance of showing the maneuver advice messages in the maneuver guidance test task. (C) desired about 30% more pre-advices ("in x m turn right") than the other users before turning right. This can be an indication for higher information speed and higher information density in China compared to Germany, for example. *POI (MD)* counts the number of points of interest shown in the navigation map display. Information density increases with the number of POI and is two times higher for (C) than for (G) or (E). *MaxOpenTasks* is a metric variable, which represents the maximum number of open tasks in the working environment (i.e. running applications and icons in the Windows ™ task bar) during the test session with the IIA data collection tool. (C) tend to work on more tasks simultaneously than (G) or (E) which can be possibly explained by the way of work planning (polychrome vs. monochrome timing, [4]) or the kind of thinking (mono-causal (sequential) vs. multi-causal (parallel) logic, [14]). *Information speed value* represents the time the maneuver advice message is visible on the screen. (C) and (G) wanted the messages to be visible about 40% longer than (E) do. *Number of Chars* contains the number of characters entered by the user during the maneuver guidance and map display test tasks in answering open questions. This is explained by the fact that the Chinese language needs considerably less characters to represent words than English or German.

The cultural interaction indicators can be visualized also when applying the IIA data analysis tool to plot "cultural HCI fingerprints" (in the style of Smith and Chang [16]) which represent the cultural differences in HCI in respect to several variables for HCI design that depend on the cultural background of the potential target group of users (Figure 2).

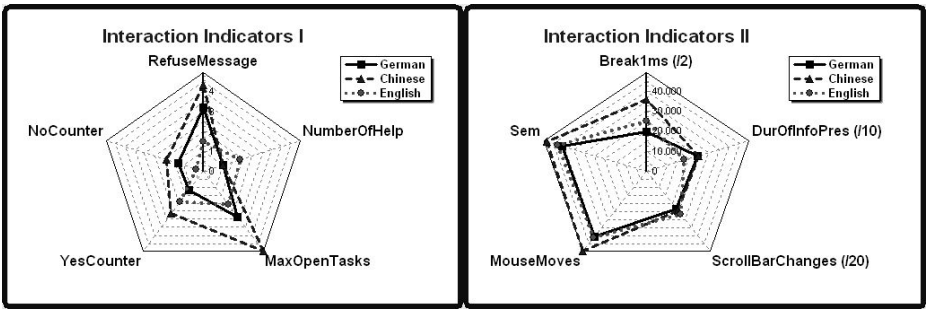


Fig. 2. Examples of "Cultural HCI Fingerprints" (different values of the cultural interaction indicators according to test languages) plot by the IIA data analysis tool

Discussion of Results: Classification Power of the Cultural Interaction Indicators. Step-by-step discrimination analysis ("Jackknife-Method") offers iterative analysis of the best discriminating cultural interaction indicators automatically out of the given set of the potential ones. The resulting discriminating rate for classifying all test users to their selected test languages simultaneously and correctly is 60% (Table 3).

Table 3. Significant Cultural Interaction Indicators

Classification results in % by statistical discriminance analysis		*Test language*	*Predicted group membership*		
Study			Chinese	German	English
1	Cross validated total: 60,80	Chinese	58,82	29,41	11,76
	Wilk's $\lambda_{1\text{-}2}$=.574**, λ_2=.855**	German	9,09	70,45	20,45
	$p_{inclusion}$=.05, $p_{exclusion}$=.1	English	29,17	25,00	45,83
2	Cross validated total: 59,90	Chinese	35,58	23,08	41,35
	Wilk's $\lambda_{1\text{-}2}$=.649**, λ_2=.850**	German	4,55	61,76	33,69
	$p_{inclusion}$=.05, $p_{exclusion}$=.1	English	6,45	29,49	64,06

The following cultural interaction indicators have been employed for the data set of both studies: interaction speed, information speed value, interaction exactness value, maneuver, maximal open tasks, number of POI, restaurants, streets, chars and uncertainty avoidance value. Applying the same method classifying the cases to two groups (instead of three at the same time), the discriminating rate increases tremendously: between German and English test language the discriminating rate goes up over 70% and between Chinese and German test language the discriminating rate is even higher than 80%.

3 Discussion: Scope, Reliability and Benefit of the IIA-Tool

The simulations of special use cases within the IIA tool can show usability problems and differences in user interaction behavior. All levels of the interaction model (physical, lexical, syntactical, semantic, pragmatic and intentional) necessary for dialog and interaction design can be analyzed with the IIA tool [7] [8]. The collected data can be quantitative (related to all test persons, e.g. the mean of a Likert scale) and qualitative (related to one single test person, e.g. an answer to an open question) [1]. The preparation of the collected data takes place mostly automatically by the IIA data collection tool that saves much time, costs and effort. The data will be stored in databases in a format, which is immediately interpretable by the IIA analysis tool that also processes possible subsequent converting or data preparation. Moreover, the collected data sets have a standard format so that anyone can perform own statistical analyses. This means that the results of this study are verifiable because they can be reproduced using the IIA tool. The statistic program SPSS, neural nets and AMOS can be directly fed with data and deployed to achieve descriptive or explanatory statistics, correlations and explorative or confirmatory factor analysis, to explore cultural differences in the user interaction as well as to find a cultural interaction

model using structural equal models. The programming language “Delphi” was used to create one single software tool, which can be provided for online download per Internet or Intranet, as well as offline per CD. To avoid downloading and interaction delays in the online version, the IIA tool is not an applet or web application but an executable program file on a file server that can be downloaded onto the local hard disk of the users worldwide. This allows the tool to measure the interaction behavior of the user during the test sessions correctly and comparably. The Delphi IDE allows transforming new HMI concepts easily and quickly into professional looking prototypes that can be tested very soon in the development process: some of the hypotheses above have been confirmed quantitatively addressing many test users online using the IIA tool within about one month (implementing the use cases as well as doing data collection and data analysis). The studies with the IIA tool comparing Chinese and German users revealed different interaction patterns according to the cultural background of the users regarding e.g. design (complex vs. simple), information density (high vs. low), menu structure (high breadth vs. high depth), personalization (high vs. low), language (symbols vs. characters) and interaction devices (no help vs. help) [7]. Furthermore, the two online studies revealed many reciprocal confirmed aspects: the high discrimination rate of over 80% and the high accordance between the cultural interaction indicators found by one-way ANOVA and Kruskal-Wallis-Test respectively. To verify the discriminating rate by a more practical method, a back propagation network has been implemented into the IIA data evaluation tool. All values of the 118 potential metric cultural interaction indicators of all data sets have been z-transformed and normalized to the range of [0;1] to be able to feed the input neurons with comparable data. Three output neurons indicate the test languages Chinese, German and English. According to the network topology and learning rate, the discriminating rate climbs up to 80% for correctly classifying the users to the used test language. Also qualitative studies, e.g. from Vöhringer-Kuhnt [17], confirm the results of the quantitative studies done with the IIA tool. All these points attest the high reliability and criteria validity of the statistical results received in this study using the IIA tool as well as the validity of the methods implemented in the IIA tool which justifies and even encourages its usage in the future for more detailed studies. Nevertheless, even if the results of these studies applying the IIA tool show that the *HCI between Chinese, German and English speaking users differs significantly* the values of some variables differ slightly between the studies. This can be explained by different test conditions or test tasks whose design has still to be optimized within the intercultural usability engineering process [11]. Additionally, studies that are more detailed must show whether changing the metrics of potential indicators (or using them in other situations, use cases or circumstances) will improve their discriminating effect and yield appropriate values accordingly. Above all, future studies have to be done to reveal relevant cultural parameters according to other user groups (e.g. elderly vs. younger people, experienced vs. beginners, female vs. male, drivers of different vehicles etc.).

4 Conclusion

The IIA tool serves to record the user’s interaction with the computer to be able to identify cultural variables like color, positioning, information density, interaction

speed, interaction patterns and their values, which enable the deduction of design rules for cross-cultural HCI design. It is effective, efficient and reasonable to use the IIA tool within the process of cross-cultural HCI design because it can be used locally and worldwide and provides quantitative comparable and reliable results whose validity and method to get them is quantitatively and qualitatively confirmed by several studies. Using the IIA tool means rapid use case design, i.e. real-time prototyping of user interfaces for different cultures as well as collecting huge amounts of valid data rapidly and easily worldwide online via internet or intranet.

5 Outlook

The IIA tool will be continually optimized based on user feedback to extend the analysis and evaluation of cultural differences in HCI by exploring cultural interaction patterns and by improving the discrimination capability of cultural interaction indicators. Questionnaires in conjunction with recording biofeedback signals (heart rate and skin response) should give controlled insights into the user preferences. The near-term objective is to develop enhanced techniques using statistical methods (factors analysis, structure equation models, cluster analysis etc.), data mining and semantic processing to extract the cultural variables and its values as well as the guidelines for cross-cultural HMI design in a more automatic way. Moreover, the method to implement new use cases will be extended (e.g. employing authoring tools or HMI description languages).

Acknowledgments. I like to thank everyone who was or is supporting the development of the IIA tool.

References

1. De la Cruz, T., Mandl, T., Womser-Hacker, C.: Cultural Dependency of Quality Perception and Web Page Evaluation Guidelines: Results from a Survey. In: Day, D., del Galdo, E., Evers, V. (eds.) Designing for Global Markets 7: Proceedings of the Seventh International Workshop on Internationalization of Products and Systems IWIPS 2005, The Netherlands, Grafisch Centrum, Amsterdam, pp. 15–27 (2005)
2. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction. Prentice Hall, London (1993)
3. Hall, E.T.: The Silent Language. Doubleday, New York (1959)
4. Hall, E.T.: Beyond Culture. Anchor Press, Garden City, NY (1976)
5. Halpin, A.W., Winer, B.J.: A factorial study of the leader behavior descriptions. Bureau of Business Research, Ohio State University. Columbus, Ohio (1957)
6. Heimgärtner, R.: Research in Progress: Towards Cross-Cultural Adaptive Human-Machine-Interaction in Automotive Navigation Systems. In: Day, D., del Galdo, E.M. (eds.) Proceedings of the Seventh International Workshop on Internationalization of Products and Systems. IWIPS 2005, The Netherlands, Grafisch Centrum, Amsterdam, pp. 7–111 (2005)

7. Heimgärtner, R.: Measuring Cultural Differences in Human Computer Interaction as Preparatory Work for Cross-Cultural Adaptability in Navigation Systems. In: Useware 2006, VDI-Bericht Nr., pp. 301–314. VDI-Verlag, Düsseldorf (2006)
8. Herczeg, M.: Software-Ergonomie. Grundlagen der Mensch- Computer-Kommunikation. Oldenburg-Verlag, Bonn (1994)
9. Hofstede, G.: Cultures and Organisations: Software of the Mind. McGraw-Hill, London (1991)
10. Hofstede, G.: The Pitfalls of Cross-National Survey Research: A reply to the Article by Spector et al. on the Psychometric Properties of the Hofstede Values Survey Module 1994. Applied Psychology: An International Review, vol. 51(1), pp. 170–173 (2002)
11. Honold, P.: Interkulturelles Usability Engineering. Eine Untersuchung zu kulturellen Einflüssen auf die Gestaltung und Nutzung technischer Produkte. VDI-Verlag, Düsseldorf (2000)
12. Marcus, A., Baumgartner, V-J.: A Practical Set of Culture Dimensions for Global User-Interface Development. In: Masoodian, M., Jones, S., Rogers, B. (eds.) APCHI 2004. LNCS, vol. 3101, pp. 252–261. Springer, Heidelberg (2004)
13. Röse, K.: Methodik zur Gestaltung interkultureller Mensch-Maschine-Systeme in der Produktionstechnik. Fortschritt-Berichte pak. Band 5. Mensch-Maschine-Interaktion, Verlag Universität Kaiserslautern, Kaiserslautern (2002)
14. Röse, K., Liu, L., Zühlke, D.: Design Issues in Mainland China: Demands for a Localized Human-Machine-Interaction Design. In: Johannsen, G. (ed.) 8th IFAC/IFIPS/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems. Preprints, Kassel, pp. 17–22 (2001)
15. Shneiderman, B.: User Interface Design. MIT Press, Cambridge (2002)
16. Smith, A., Chang, Y.: Quantifying Hofstede and Developing Cultural Fingerprints for Website Acceptability. In: Evers, V., Röse, K., Honold, P., Coronado, J., Day, D. (eds.) Proceedings of the Fifth International Workshop on Internationalization of Products and Systems. IWIPS 2003, Germany, Berlin, 17–19 July 2003. University of Kaiserslautern, Kaiserslautern, pp. 89–102 (2003)
17. Vöhringer-Kuhnt, T.: Asiatische vs. europäische HMI Lösungen von Fahrerinformations systemen. In: Useware 2006, VDI-Bericht Nr. 1946, pp. 279–287. VDI-Verlag, Düsseldorf (2006)

Locating Culture in HCI with Information Kiosks and Social Networks

Tom Hope, Masahiro Hamasaki, Keisuke Ishida, Noriyuki Fujimura, Yoshiyuki Nakamura, and Takuichi Nishimura

Information Technology Research Institute, National Institute of Advanced Industrial Science and Technology, 2-41-6 Aomi, Koto-ku, Tokyo, Japan
{tom-hope,masahiro.hamasaki,ksk-ishida,nori.fujimura, nakamura-y,takuichi.nishimura}@aist.go.jp

Abstract. Concepts of 'culture' are often invoked in analysis of human-computer interaction, notably in attempts to refine or adapt systems to differing cultural contexts, such as in the process of internationalization or in creating systems and processes that can adapt to user's cultures. This paper takes ethnographic research in this area to the study of culture in HCI to address culture as a problematic unit of analysis. It does this via qualitative video-based analysis of user's interactions with information kiosks at international conferences. The paper argues that culture must be understood as contingent and nationality may not be the most important indicator in multi-national co-located settings.

1 Introduction

The importance of examining culture in HCI has arrived primarily through the recognition in the field of marketing that certain products are more likely to be accepted if modified to fit with 'indigenous' cultural values [12]. One very large facet of the contemporary computer industry is the continued expansion of markets in developing countries and innovation in already established markets. Computer systems, complete with designed interfaces are transferred between cultures and used in many contexts by people with differing characteristics. Concepts of 'culture' are often invoked in analysis of human-computer interaction, notably in attempts to refine or adapt systems to differing cultural contexts, such as in the process of internationalization [2], or in creating systems and processes that can adapt to user's cultures [3]. A research tradition of studying users in their place of work, building from a foundation in ethnomethodology and conversation analysis continues to explore differing uses of computers and associated technologies within their situated 'cultures' [17]. Work in cognitive science [9] and analysis of conversation structure and practice [15] has questioned understandings of cognition, perception and 'understanding' as being located within the individual. It is clear that the importance of considering culture in HCI shows no signs of decreasing, as globalization continues.

This paper uses analysis of users of a conference support system to 'locate' culture in the HCI and asks the question, "when is culture important?" This is then followed by a description of our conference support system, UbiCoAssist. Following a brief

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 99–107, 2007.
© Springer-Verlag Berlin Heidelberg 2007

explanation of the research methodology, the paper presents analysis of users' interactions with the kiosk interface of UbiCoAssist. The focus in this paper is on the interactions of groups or 'communities' [10] of users at the kiosks. The results of the study are then presented, which raise issues of what we can take as 'cultural' behavior. This is followed by a brief note of how culture may be kept in analysis of human-computer interaction.

2 Related Work: Culture in HCI

Early studies in interaction with computers did not place very much emphasis on the cultures of the users under study, but in recent years, perhaps partly due to the expansion of the Internet, cross-cultural studies in HCI have become popular. Much of this related work, not surprisingly, focuses on website design and the need to develop interfaces that are culturally appropriate [18, 13]. It has also been argued that cultural bias exists at the cognitive level of the designers of sites, which can affect the uptake of products and information [5]. Clearly, in a rapidly globalizing world, this presents a problem.

The typical response to the issue of cultural diversity is to divide cultures into diverse groups in order to design for particular cultural traits and a great deal of research has been done on the meaning of colors and other aspects of culture such as metaphor [2]. One of the most extensive and often used studies is that done by Hofstede [8], based on his work at IBM in the 1970s. In this research and that which follows it, culture is essentially a set of dimensions that vary according to nationality. Following a cognitive psychological model, culture is viewed as "the collective programming of the mind", and is understood as statistically measurable and able to be visualized in graphic form. Thus, by applying the models, it is believed that effective localization of interfaces can be developed. Various methods have been proposed as to how to best gather the data for cross-cultural interface design, such as noting cultural 'markers' in observation [12]. From the subsequent models, a process of interface development can begin.

There is another prominent meaning of culture in HCI, however, which reflects more recent concerns for localized context and setting, rather than using nationality as a primary indicator. Authors have examined the located and contingent nature of interaction with computers [17], arguing that users are not 'cultural dopes' [6], but are active in creating the culture as they live it.

The practice of using 'cultural probes' in interaction design, originating from Gaver et al [7] illustrates how culture can be understood as more local, using members on the inside of the culture to record practices. Indeed, from this viewpoint, culture is less to do with large-scale histories and instead related to localized established practices. Much related methodological work is in this tradition, particularly those that aim to place new systems in homes or observe directly people's behavior inside their homes [4, 1]. This latter approach—that of culture as directly observable and located in settings informs the research for this paper. Technologically, the most related research is Okamoto et al [14], who developed a system to display user's cultural competencies on a shared screen. The system provides information about other users' acquaintances, language ability and length of

time in the culture. Though primarily a system reflecting nationality and linguistic community, the connection to our system is strong in that users can see information about others, and is designed to be used in international contexts.

3 System Design and Implementation

In this section we present an overview of the system, its basic architecture and user interface. The system we developed is called UbiCoAssist, a shortened name for "Ubiquitous Community Assistance".

3.1 UbiCoAssist System Overview

UbiCoAssist is a system that is to be used at conferences and similar events to enable the participants of those events to create links with each other and maintain their community before, during, and after the event. The system consists of three parts: Software which mines the world wide web in order to generate social networks of participants, a web-browser interface where users can view and modify their network, and an on-site information kiosk-based interface where users can view, modify and join each others networks using RFID IC cards (Fig.1).

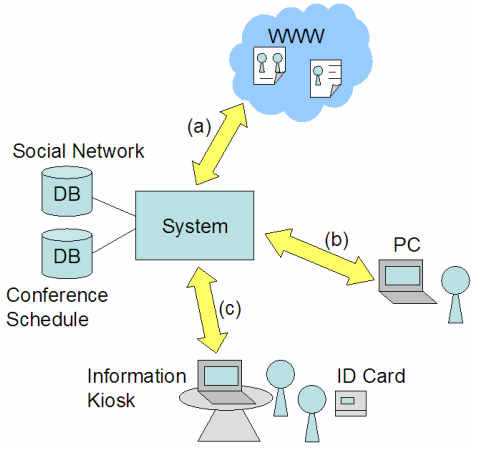


Fig. 1. System overview showing the three parts to the system: (*a*) web-mining gathers information from the world-wide-web, (*b*) a web interface for users and (*c*) a kiosk-based interface used with ID cards

When users log into the system using their own PC or one of the kiosks on site, they are presented with a homepage ('Mypage'), which contains information about their network of friends and colleagues at the event. It also includes a visualized social network diagram (Fig. 4).

Information kiosks onsite provide the ability for users to login simultaneously with others, thereby creating network links if not already existant and viewing joint social networks on the kiosk display.

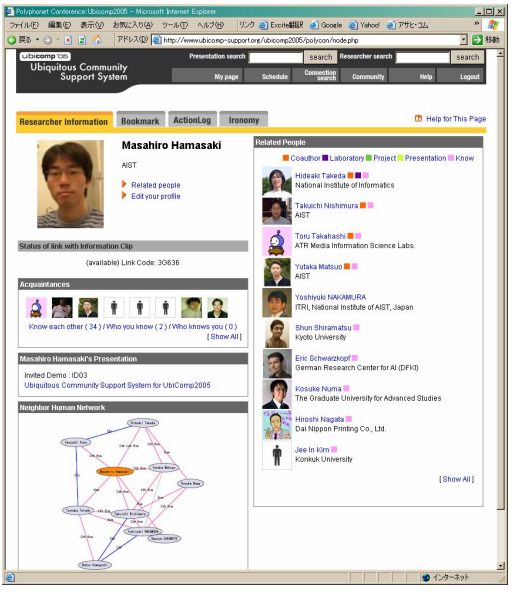


Fig. 2. 'Mypage' is the user's homepage, from which they can browse and modify their network, the event schedule and previous actions

IUbiCoAssist extracts social network information in three ways. The web-mining is automatic upon participators registration to the conference (or to the system). A 'know-link' is created when users select to add other participants to their network via the browser interface. Finally, using an RFID ID card that is given to users upon arrival at the kiosk, they can jointly make 'touch-links' by placing their cards on readers in front of the displays.



Fig. 3. Users gathered and gesturing at an information kiosk

The system has been employed at several national and international conferences inside Japan and demonstrated overseas. The data for this paper is taken from the version of UbiCoAssist used at an international computing conference in Tokyo in 2005.

4 Methodology

Video cameras were used to capture user interaction at the conference. The data was gathered over the entire period of the event, in order to allow observation of users beginning to use the system and potentially gain competency of its use. The analysis of this qualitative data takes as its foundation earlier work in the social aspects of HCI and human interaction in general, including Suchman [17], Garfinkel [6] and Sacks [15]. An analysis of the work of 'doing' community behaviour at the kiosks has been presented in Hope [10].

As we have seen, culture in HCI is frequently utilized to mean 'nationality'. Therefore, we decided to examine user login patterns according to their country of residence. As users had not been asked to enter their nationality when registering to the system, we achieved a working understanding by separating the data according to email address domain. In order to register to the system so that the web-mining would function appropriately, all users had to register with their email address and enter their affiliation. Using the domain (e.g. .jp, .au, .kr, being Japan, Austraila and Korea, respectively) we were able to determine the location of the institute—for it was usually an institute's email address that they used—and therefore the country where they are most likely to live and work. This method clearly has some problems—users may not register with their usual or present work address, or the address that they use may not be related to the actual location where they live. However, within these limitations it was possible to see patterns in behavior.

Using the log of system usage, along with fieldnotes taken at the conference, we were able to select potential areas of cultural difference, whereupon we used the video data to observe closely the interactions that occurred.

5 User Study

Examining the video data and fieldnotes highlighted seemingly simple, but extremely common issues experienced by many users of the system. These are related to the usage of the RFID cards during login, which in turn builds on cultural foundations of social status. We explore these issues in this section.

5.1 Patterns of Login Error

UbiCoAssist has been designed to be relatively self-explanatory. While the interface language used at the conference was English, this was also the language of the conference itself, so, barring issues of metaphors or idioms, of which there were none, the instructions of how to use the system were clear. At the point of registering an RFID card to one's name, the system presented instructions to be followed on the display. Entering email addresses caused some expected difficulties due to an unfamiliar keyboard layout, but it was the act of using the RFID card itself, which caused problems. To log in, a user must place the RFID card onto the sensor and leave it on the sensor for the entire period when they desire to be logged in. Removing the card consequently logs the user out of the system. This is to prevent users from leaving the kiosk while still logged in, thereby risking their private data being used by

others. The problem was that some users did not recognize this fact and so, in their communication with the system, attempted to log in by swiping the card over the reader or touching it and removing it immediately. Presently computers, or at least UbiCoAssist at this stage, are relatively unable to recognize when there is a breakdown in communication between the user and itself. In human-human interaction, when breakdown occurs, a process of 'repairing' the 'error' begins, negotiated by the actors in the interaction [11]. In the case of UbiCoAssist, the user error generates a rapid login and logout, leaving the kiosk display showing the pre-login screen, which must then be interpreted by the user. We examined the data with the hypothesis that we would see different patterns of this user error according to nationality. Fig. 4. shows the ratio of rapid logging in and out (under 5 seconds) to correct viewing of the system (over 5 seconds) by domain. We have removed those domains where the number of individuals was too few as to make the results meaningful.

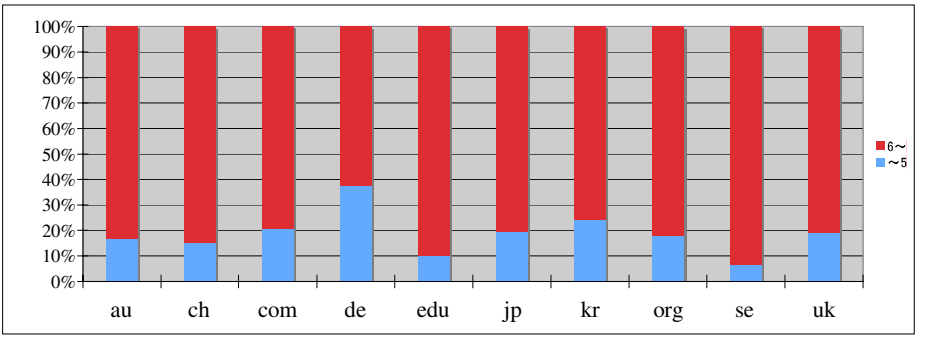


Fig. 4. Ratio of user error to correct usage, with user error displayed at the bottom and proper viewing at the top

As we can see, variance between domains is not great between Australia (*au*), Japan (*jp*), Korea (*kr*) and the United Kingdom (*uk*). The .com, .edu, and .org, domains are also at a similar level, although it would be wrong to assume that these denote users from the United States only. It should be noted that not all members in each group would make a login error at their first usage of the system, but this was very common.

Using this data, it was possible to see which people in each group were making the errors and then, by using a video timestamp, view these interactions on the recorded video data. The aim in this was primarily to see if there were observable cultural dimensions that affected their behavior. We initially assumed that users' login errors were due to unfamiliarity with RFID cards. Our expectation prior to analysis was, for example, that those participants from cities that have popular RFID/IC card based transportation systems would have fewer errors than those without them.

Sweden (*se*), with the lowest ratio of errors and Germany (*de*) with the highest also had fewer members than the others, which may account for their results. As the conference occurred in East Asia, the majority of participants were from Japan or Korea. Only a small number were from Europe. While we recognize the limitations of this type of analysis it is relatively clear that the differences are not great enough to

warrant a statement that nationality greatly affects users' understanding of how to use the RFID cards. How then to account for the difficulties users experienced when they did occur?

Essentially, the user, regardless of whichever nationality they were, and being given no other signal from the system, often assumed that this was a system error after the first attempt. They would then attempt to log in again and learn the correct way to use the interface. The event, being a computing conference, provided a setting whereby system errors and glitches would be expected, as demonstrative systems and prototypes were on display. In this case, therefore, the culture of computer science made available a strong context for user understanding of errors.

How did other users avoid making initial login errors? The primary process was by viewing their colleagues or other friends in their networks logging into the kiosks. Those instances when users logged in correctly at their first attempt, were often when they were in the company of either staff members or other users now familiar with the system. For this reason, it is clear that the order of access to the kiosks is important, as is discussed below.

5.2 Culture and Access

Those around the kiosks socially ordered the access to UbiCoAssist. For example, in cases where a user made a login error while in the presence of a member of UbiCoAssist staff, the staff member in a position of authority would have priority of access to the system from which point they were able to log in with their own cards. This act was then swiftly learnt by the user who had just made the mistake and any others watching. Often there was no direct reference to the users error, but simply a demonstration by the staff member. In situations when there were no staff members present, more experienced members of the user's network who were at the kiosk would help the user to log in correctly.

There did appear to be a significant cultural difference in the way that some Japanese and Korean groups organized their access to the kiosk card readers. It has been well documented that the education systems and prominent cultural values of these countries develop systems of seniority among citizens [16]. When faced with the necessity to have one of the members around a kiosk log in first, some ordering based on the apparent seniority of members occurred. In many cases, the older members appeared to have priority of access and younger members would access once the senior user had logged out. By observing how to log in, other uses could then view their social networks successfully.

It is evident from the above that national or ethnic culture can affect the order in which access to card readers occurs. But in examining user errors qualitatively, it appears that this cannot be taken on face value as affecting all subsequent interactions; more localized practices come into play after the initial access, dependent on who else is present. Errors became understood in terms of computer engineering cultures in some contexts. The access to readers became more complex when members of staff were there to intervene. We can begin to understand these relatively transient, often international, communities of learning as cultures in themselves, with their own processes and practices to be explored.

Where culture is important, then, is in the initial ordering of access to the card readers at the kiosks. In other words, culture affects the order of who will make the errors, rather than the nature of the mistakes themselves (for the log-in problems were common to all nationalities).

6 Implications and Further Work

One of the distinct characteristics of our system is that it is not fixed to one physical setting; in fact it is a system that only exists for the length of a shared public event. This is significant, as it presents challenges for an understanding of 'located' culture, whether based on nationality or a culture of the home or workplace. With UbiCoAssist, the location changes with each event, as do the users. As is suggested in our analysis, in international contexts the uncritical use of static models of national culture may not be ideal. Cognitive models of culture are particularly useful for the design of interfaces in relatively stable settings, such as in cross-national communication systems, but in real-world co-located settings, such as those where UbiCoAssist is utilized, interfaces can and do become resources from which national and more local cultural practices emerge together. In future work, these more local, contingent practices should be explored to see how different aspects of the interface and system affect interaction between national and international users.

Acknowledgements. This research has been supported by NEDO (New Energy and Industrial Technology Development Organization) as the project ID of 04A11502a.

References

1. Aipperspach, R., Rattenbury, T., Woodruff, A., Canny, J.: A Quantitative Method for Revealing and Comparing Places in the Home. In: Proc. Ubicomp 2006, Orange County, CA
2. Bourges-Waldegg, P., Scrivener, A.R.: Meaning; the Central Issue in Cross-Cultural HCI Design. Interacting with Computers Special Issue 9(3), 287–309 (1998)
3. Carrara, P., Fogli, D., Fresta, G., Mussio, P.: Toward overcoming culture, skill and Situation hurdles in human-computer interaction. Universal Access in Information Society Journal (UAIS), special issue on "Continuous Interaction in Future Computing Systems" 1, 288–304 (2002)
4. Cheverst K, Clarke K, Dewsbury G, Rouncefield M, Sommerville I, Blythe M, Baxter G, Wright, P.: Gathering requirements for Inclusive Design. 2nd BCS HCI Group workshop on Culture and HCI: Bridging Cultural and Digital Divides. - June 18th 2003, University of Greenwich, England. In: Gunter, K., Smith, A., French, T.(eds.) (2003) Proceedings of 2nd BCS HCI Workshop on Culture and HCI: Bridging Cultural and Digital Divides, pp. 65–71. University of Greenwich [Reviewed Workshop Proceedings, 2003] ISBN 1 86166 191 6
5. Faiola, A., Matei, S.A.: Cultural cognitive style and web design: Beyond a behavioral inquiry into computer-mediated communication. Journal of Computer-Mediated Communication, vol. 11(1), article 18 (2005) http://jcmc.indiana.edu/vol11/issue1/faiola.html

6. Garfinkel, H.: Studies in Ethnomethodology. Polity Press, Cambridge (1996)
7. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural probes. Interactions 6(1), (January 1999), 21–29. DOI= http://doi.acm.org/10.1145/291224.291235
8. Hofstede, G.: Culture's consequences: International differences in work-related values. Sage, Newbury Park, CA (1980)
9. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput.-Hum. Interact. 7(2), 174–196 (June 2000) DOI= http://doi.acm.org/10.1145/353485.353487
10. Hope, T., Hamasaki, M., Matsuo, Y., Nakamura, Y., Fujimura, N., Nishimura, T.: Doing Community: Co-construction of Meaning and Use with Interactive Information Kiosks. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 387–403. Springer, Heidelberg (2006)
11. Hutchby, I.: Conversation and Technology: From the Telephone to the Internet. Polity Press, Cambridge (2001)
12. Jagne, J., Smith-Atakan, A.S.: 2006. Cross-cultural interface design strategy. Univers. Access Inf. Soc. vol. 5(3), pp. 299–305 (October 2006) DOI=http://dx.doi.org/10.1007/s10209-006-0048-6
13. Jagne, J., Serengul, S., Curzon, P., Fields, B.: Integrating social and cultural variances into international eCommerce interface design, vol. 2, Proceedings of HCI 2006 (2006)
14. Okamoto, M., Isbister, K., Nakanishi, H., Ishida, T.: Supporting Cross-Cultural Communication with a Large-Screen System. New. Generation Computing 20(2), 165–185 (2002)
15. Sacks, H.: Lectures on Conversation. Blackwell Publishing, Oxford (1995)
16. Sugimoto, Y.: An Introduction to Japanese Society. Cambridge University Press, Cambridge (1997)
17. Suchman, L.A.: Plans and Situated Actions: the Problem of Human-Machine Communication. Cambridge University Press, Cambridge (1987)
18. Sun, H.: Building a culturally-competent corporate web site: an exploratory study of cultural markers in multilingual web design. In: Proceedings of the 19th Annual international Conference on Computer Documentation, Sante Fe, New Mexico, USA, October 21 - 24, 2001, pp. 95–102. ACM Press, New York, NY (2001) DOI=http://doi.acm.org/10.1145/501516.501536

HCI and SE – The Cultures of the Professions

Joshi Anirudha

Industrial Design Centre, IIT Bombay, Mumbai 400076, India
anirudha@iitb.ac.in

Abstract. The author reviewed and participated in several exemplar industry projects from the Indian IT industry to study the integration of human-computer interaction (HCI) design into software development by Indian software vendors. While several problems occurred because HCI skills were either not used, or were not used early enough in a project, or when the HCI professional lacked process support to carry out all HCI activities in the project, at least some of the problems occurred because of the cultural differences between the professions of designers and engineers. In the one case where HCI professionals were indeed used early and with a multi-disciplinary team, the results were positive. The case studies point to a greater need to integrate HCI into existing software engineering process models with commonly accepted roles, activities and deliverables leading to mutual respect between professions.

Keywords: HCI and SE integration.

1 Introduction

The processes of human-computer interaction (HCI) and software engineering (SE) should, in theory, affect each other deeply as they are primarily concerned with creation of common artefacts. Traditional SE literature has been found to have several lacunae from HCI design perspective [1]. The chief among these have been that many HCI decisions happen during requirements specification and that HCI design is viewed as only skin-deep activity by SE literature. Elsewhere [2], the author had argued that educational institutes offering programs in HCI will do well by becoming more interdisciplinary in their educational approach.

Software development has in the past been compared to making of films [2]. Like software, films are made by groups of multidisciplinary, creative professionals (director, writer, cinematographer, editor, sound technician, musician, actor etc.). The different disciplines in the film world seem to be mutually respectful, well-coordinated and woven together in a common 'professional' culture.

By contrast, there exist major gaps between HCI and SE, both in academic institutions and in the industrial practice [3]. Deliverables of one group are not evidently useful to the other, hampering workflows. There is an apparent disconnect between the priorities of the two groups. These gaps lead to communication and coordination problems, duplication of effort, compromises in the process and eventually the quality of the output.

Here we focus on the differences arising between the culture of professions of HCI designers and software engineers in practice. We hope to identify these through some

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 108–112, 2007.
© Springer-Verlag Berlin Heidelberg 2007

industrial case studies from the Indian IT industry and identify some lessons we can learn from them.

2 Case Studies

The author reviewed as case studies real-life projects from the Indian IT industry to understand how the work of software development actually happens in practice in the process-conscious Indian IT industry. The author participated in some projects in the role of a HCI designer or usability evaluator. Additionally, professionals in the industry were requested to contribute live case studies. The questions we were trying to answer were:

- How and when do HCI design decisions happen in SE process?
- What role do HCI design professionals play? How do they influence the SE process?
- What are the common objects between HCI and SE process?
- Are the concerns [1] about HCI in SE literature handled? How?

Many contributors were concerned about confidentiality and information security policies within their organization and their contractual agreements with the clients. In spite of these difficulties, seven case studies could be analyzed. The case studies are anonymous to protect the identity of the concerned on their request and all confidential information was removed before the case study was used in this document.

While selecting the case studies from within these constraints, we have tried to make sure that they represent typical cases rather than exceptions. We have also tried to represent a variety of organizations in terms of their size and usability maturity.

2.1 A Last Straw Project

A UK-based company S has a legacy software product for universities. S was migrating this product to a .Net platform in a phased manner. S outsourced part of this phased migration activity to a software development company T in India. T hired a design consulting company D to help T to do a usability evaluation of the product at a time when the product was nearing its delivery. Heuristic evaluation was done by five evaluators, several usability problems were identified and design solutions suggested. However, the product got delivered with several known usability problems as it was too close to the delivery date to implement many changes. Some of the usability problems could have been easily fixed had they been identified sooner.

2.2 HCI Design Without HCI Designers

A large, CMMi level 5 company was doing a ‘rewrite from scratch’ of a large legacy system with over 130 use cases to an extranet application for an insurance client. There were no HCI professionals involved in this project. An inspection of the requirements documents revealed that a good number of premature (and poor) HCI

design decisions were already taken and specified in the use case documentation. No apparent HCI design process was followed to make these decisions.

2.3 Users Reject Unusable Software

A finance company in the US, X upgraded one of their desktop-based applications by porting it to J2EE technologies. X awarded this project to N, a large IT organization in India. Usability professionals from N were involved, but they lacked sufficient process support and their estimates went haywire. As a result they could not do any contextual user studies. The delivered software lacked an important need of users – keyboard shortcuts, which was noticed during user acceptance tests. The task completion time for a frequent task went from 5 minutes to 20 minutes and users rejected the product outright. N was obliged to carry out the changes at its cost.

2.4 Usability Group in Project Conceptualization

A mid-sized Indian company U was migrating an application from laptop to a mobile phone based application for use by thousands of field engineers of a UK based telecom company. U set out a team comprising of software engineers, business analysts and HCI designers to identify a business case for the project. The multidisciplinary team successfully created a proof of concept and justified the business case that went into development.

3 Lessons About the Cultures of Professions

The following were the lessons learnt about the cultures of professions from these and other case studies:

- "Design during requirements!"
  Use cases over-specify HCI design, particularly in the context of outsource software development (perhaps to exercise better control on the output). Getting these HCI design decisions right is very crucial. A wrong specification in the use case can have a ripple effect in several activities of software development and may lead to avoidable iterations. Hence HCI design and usability evaluation inputs are needed early in a software project – certainly before freezing UI design, and possibly before freezing on requirements. Late inputs have little impact and are expensive, early inputs have significant impact and are cheaper. However, because of the terms 'design' and 'evaluation' associated with HCI activities, traditional software engineers are averse to accepting and including these activities during the requirements specification phase.
- "A process is not a formula"
  The growth-oriented, scaling-up culture of Indian IT organizations has the underlying assumption that any person can do any activity and everyone is dispensable. When HCI skills are not available, HCI design decisions are taken by the available people. When skilled HCI professionals are available, and when they do a good job, managers tend to worry about the 'un-repeatability' of the work they do. No matter how process-driven one is, HCI design is person-dependant

activity. In other words, given the same process, one designer can often do a better job than another designer. While a well-defined design process can streamline activities and improve the quality of the outcome, creativity of the individual designer plays an important role. Companies tend to equate HCI skills to other skills (such as .Net or Java) and expect high level of compatibility.

- "Client centred or user centred?"
  Several HCI designers have a user-centred design culture. On the other hand, offshore software development companies seem to have a client-centred culture, particularly for HCI design. In typical offshore projects, requirements documents are created onsite, in collaboration with the client. The client is often called upon to specify many HCI designs in the guise of specifying requirements so that the off-shore development can happen smoothly. Similarly, if a product tests well during an acceptance test by the client, it is considered to be usable by end-users as well. While the client is an important stakeholder in the design process, client-centred approaches don't by themselves lead to optimal design.
- "It is important to approach HCI design holistically"
  Engineers are typically trained in reductionist, deductive thinking patterns. For example, SE aims at minimizing coupling between modules and maximizing cohesion within modules. This allows companies to break up large projects into smaller modules, which are developed by smaller teams and with minimal interaction across teams. It also allows companies to develop modules sequentially. Unfortunately, same approach can't be used in HCI design. Inconsistencies creep in a product user interface if it is designed by different people at different times. Even if development of a project is scheduled to happen piecemeal, early HCI design inputs can develop the bigger picture and a rollout plan. This leads to better software design as well as a better UI design. However, such approaches are usually not taken during software development. Several engineers may accept holistic approaches in principle, but these are often shot down in practice due to time and budget constraints.

4 Conclusion and Further Work

Though there were several limitations in these case studies (such as their number and the details that one can study), some trends emerge. The return on investment of usability and HCI has been established extensively. However, these case studies show the need for better integrating HCI activities into the software development process. SE techniques were developed to create multi-version software by multi-person teams. They need to be expanded to include multi-disciplinary teams. Further, it is important to integrate the cultures of the HCI and SE professions and inculcate a feeling of mutual respect among professionals.

Going forward, we plan to propose process models that integrate HCI with SE processes. In this, we propose to clearly define roles, activities and deliverables in a common multi-disciplinary process framework that is acceptable to both professions.

Acknowledgments. These case studies were studied as a part of an ongoing PhD research by the author. For reasons of confidentiality, it is necessary to hide the identity of the contributors of case studies. All the same, the author is grateful to them

for their effort, data and time. He is also thankful to his guide Prof. NL Sarda of IIT Bombay for his valuable guidance and suggestions.

References

1. Joshi, A.: HCI in SE Process Literature, Indo-Dan HCI Research Symposium, IIT Guwahati (2006)
2. Joshi, A.: Education of Interaction Design - an Interdisciplinary Approach, Design Education - Tradition and Modernity, Ahmedabad (2005)
3. Bridging the SE & HCI Communities (accessed March 24 2005) http://www.se-hci.org/bridging/index.html

Development of Integrated Analysis System and Tool of Perception, Recognition, and Behavior for Web Usability Test: With Emphasis on Eye-Tracking, Mouse-Tracking, and Retrospective Think Aloud

Byungjoo Kim, Ying Dong, Sungjin Kim, and Kun-Pyo Lee

Abstract. Recent researches reveal effort to observe user's experience from user's point of view in order to estimate usability of a web site. Eye-tracking and mouse-tracking to record and analysis what user sees and how user acts can be proper examples. However, although eye-tracking and mouse-tracking are used practically, not only difficult to find the case that uses both, but also the case is rare that considers what user is thinking. Hence, this paper introduced EMT System that tracks eye and mouse, and records user's thinking. And for applying EMT system, this paper developed EMT Tool, which helps a researcher to do usability test by recording the user's experience, and reproducing it visually. EMT Tool is consist of EMT Tracker which is responsible for observing and collecting user's experience and EMT Analyzer synthesizing and analyzing data from EMT Tracker.

Keywords: Web Usability Test, Integrated Analysis System, Eye-Tracking, Mouse-Tracking, Retrospective Think Aloud.

1 Introduction

Since World Wide Web (WWW) was firstly introduced by Tim Berners-Lee in 1989, it has been widely used by public nowadays, and influence the society and culture in many aspects. Following web popularization, web business is growing rapidly. As the competitions among global companies were deepen, the concern of web usability was more and more emphasized. Jacob Neilson said: "web users undergo the usability before using a web or deciding a latent purchasing",[1] which is to emphasis web usability through realistic problems in order to get web business success.

People began to consider the importance of web usability. Many researches which relate to usability evaluation methods are actively carried out in order to improve the web usability. Various usability evaluation methods such as user questionnaire, Heuristics evaluation, Usability test, and Card sorting are widely used for web usability evaluation. But, such usability evaluation methods were originally developed for products and software, some problems including subjective evaluation from evaluators' point of view, difficulties of user observation, and limitations of grasping users' needs come out if apply those methods to web usability evaluation. Moreover, as the asynchronous communication is expending with the web2.0's appearance, the

[1] Jacob Nielsen, Designing Web Usability, New Riders Publishing, 2000.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 113–121, 2007.
© Springer-Verlag Berlin Heidelberg 2007

observing method of users' interaction through Web Server's Data Logging is loosing the function of evaluating web usability.

This study tries to overcome the limitations of existing usability evaluation, and aims to develop analysis methods based on users' experiences instead of the web usability evaluator's perspectives.

2 Web Usability Methods and User Experience Review

2.1 Classification of Web Users' Experience

Existing web usability evaluation methods have not satisfied the sub-conscious, tacit, emotional, and qualitative characteristics of the web observation and analysis required for the web usability evaluation. And those have been used on the evaluators' point of view. So that above problems comes out. It is very important to observe users' experience in users' perspectives in order to evaluate web usability.

As mentioned above, it is necessary to understand users as the core of the usability to observe web users' experience at firsthand. In the web environment, users could have various and complex experiences in a short time. So it is proper to apply observation method according to different users' experience sections.

By comparing and synthesizing Norman's seven Stage of Action, Trumbo's Spatial Environment in Multimedia Design, and Kantowitz's Human Information Processing Model, three sections such as perception, recognition, and behavior of users' experience can be drew, as the table 1 shows

Table 1. Comparison of users' experience categories

	Perception	Recognition	Behavior
Norman's Seven Stage of Action	Perceiving the State of the World	Interpreting the Perception, Evaluation of Interpretations, Goals, Intention to Act, Sequence of Actions	Execution of the Action Sequence
Trumbo's Spatial Environment in Multimedia Design	Perceptual Space	Conceptual Space	Behavioral Space
Kntowitz's Human Information Processing Model	Perceptual Stage	Central Processing	Action Stage

Users' perception is the procedure of knowing and long-time memorizing the outer world stimulus through the sensible organs. The recognition is the procedure of setting the action plan for realizing aims by knowing and evaluating the inputted information. Behavior is the procedure of expressing the real executions by following the action plan. Users' experience sections such as perception, recognition, and behavior are based on users' recognition models. So it can be used not only on the interaction between physical objects like products and users, and also the interaction between web and users.

2.2 Observation Methods for Each Specific Experience

By following web users' experience section of perception, recognition, and behavior as mentioned above, using different user observation method according to each specific experience is effective on web usability evaluation.

Eye tracking is suitable for web users' perception experience observation. While web users staring at computer interface, web contents such as text, graphic image, and various multimedia can be perceived. In the graphic user interface of web environment, user's visual sense takes great proportion in perceiving. The basic principle of eye tracking is to analyzing eye movement by measuring eyeballs' movement such as fixation and searching.

In order to observe web users' recognition experience, 'Think Aloud' is an existing usability evaluation method which lets users to speak out what they think during the web usability evaluation process. Because users can explain their perception process if they speak out,[2] evaluator can observe users' recognition process through their words. Usually, 'Think Aloud' is used to let users to speak out their thought when they are performing tasks. But, at the same time, 'Think Aloud' results users' unnatural behavior and less concentration during the performance. It is not suitable for eye tracking method which is used for users' perception observation. Another 'Think Aloud' method, which is called 'Retrospective Think Aloud', asks users to speak according their memories after they finish performing those tasks. The possibiliy of reliance problem exists because users provide untruthful description in order to make them sounds more reasonable. But, Zhiwei Guan[3] proofed that the retrospective think aloud method has satisfied reliability and validity. The retrospective think aloud was used as web user's recognition experience observation method in this study.

Mouse tracking is suitable for web users' behavior experience observation. In most interaction of GUI web interface, cursor's movement can be captured.[4] Compare to Video Ethnography which is suitable for observing users' movement in the physical world through the device such as video camcorder, mouse tracking is useful for observing web users' movement which appears on the monitor's screen background.

3 EMT System: Integrated Evaluation System of Perception/ Recognition/Behavior

3.1 Introduction of EMT System

Web usability evaluation aims to understand users' latent think or interaction pattern by observing web users' usage. In order to observe web users' experience from all sides, this paper proposed observation method for users' specific experience and "EMT (Eye, Mouse, Think) System; Perception/Recognition/Behavior integrated

[2] Jeeny Preece, Human Computer Interaction, Addison-Wesley, 1994.

[3] Zhiwei Guan, Shirley Lee, Elisabeth Cuddihy, Judith Ramey, The validity of the stimulated retrospective think-aloud method as measured by eye tracking, Proceeding of CHI2006, 2006.

[4] ChangMin Park, A Study on the Visulization & Analysis of User Interaction on the WWW, Master Thesis at KAIST, 2001.

evaluation system" which can analyze web usability problems. EMT System which includes eye tracking, mouse tracking, and retrospective think aloud which are parallel used for observing users' specific experience is a systematic method, which synthetically analyzes observed users' perception process from specific experience.

3.2 Constitution and Progress of EMT System

To observe meaningful web usability problems from users, EMT System is based on the Usability Test which can get much detail data from users.

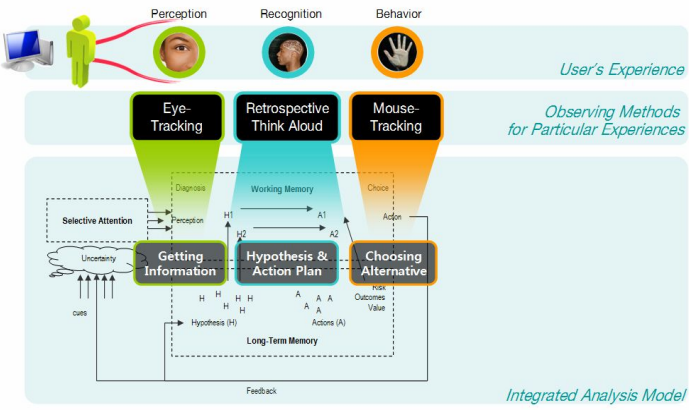


Fig. 1. Constitution of System

In the usability test as showed in figure 1, web users' experiences were observed when they were performing the given tasks. The perception of each user was observed by eye tracking, the recognition was by retrospective think aloud method, and the action was by mouse tracking. Each specific experience data of observed users were analyzed according to the integrated analysis framework.

The usability evaluation experiment was carried out according to the usability test matrix in which the evaluation system was referred is showed in the figure2 bellow.

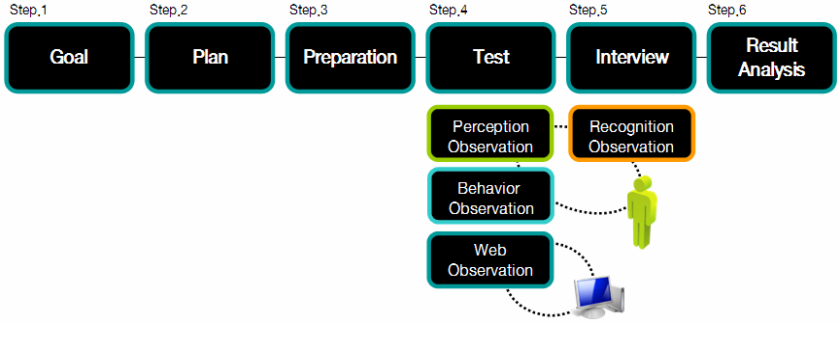


Fig. 2. Progress of EMT System

3.3 Analysis Frameworks

The EMT System analyzes the collected web users' specific experience observing data synthetically through human's information processing model of decision making. During performing the task in web users' evaluation experiment, user makes decision among many alternatives. This process related analysis can be used for analyzing web usability problems. Decision making basically contains 3 phases: getting information, making hypotheses, choosing plan and execution. According to the decision making processing model, users make many alternatives or hypotheses by using a few clues from external stimuli and then choose one of the alternatives for the action plan based on users' memory and finally carry out.[5] Those 3 phases of the information processing model of decision making, as showed above, can be applied to and be analyzed with Perception/Recognition/ Behavior phase of human experience.

Table 2. User's Particular Experiences and Steps of Decision Making

User's Experience	Step of Decision Making	Description
Perception	getting information	Getting information and clues from external stimuli
Recognition	making hypotheses	Making hypothesis and action plan based on long-term memory
Behavior	choosing plan and execution	Executing the chosen action plan

In the EMT System, the process of users' perception can be understood as the process of getting information in the decision making model. Through eye-tracking data, users' picked information and clues from web page can be analyzed. Next, the process of recognition can be understood as the process of making hypotheses. With the description from retrospective think aloud, it is possible to analyze how users set up the hypotheses and action plans for the task, under exposure to visual stimulation or information from the web page. Finally, the process of behavior can be understood as the process of choosing alternatives. By following the hypotheses and action plans in users' recognition process, how to execute specifically can be analyzed through mouse-tracking data.

The information processing model of decision making is different depending on users' different degree of expertness. Because the novice and expert have differences in terms of information processing, task execution path, background knowledge and experience, they would act differently in each phrase of getting information, making hypothesis, and choosing alternatives. Based on Gordon's skill-rule-knowledge model, the EMT System divides users' degree of expertness into analytical behavior, intuitive behavior, and automatic behavior.

As showed in table 3, in the process of performing the given tasks, the proportion of each specific experience is different according to different degree of expertness. Because the novice needs to process much information in the process of Perception/Recognition/Behavior, the observed data from each experience are also huge. Especially, the data from the complex thinking process are huge. On the contrary, the intermediate level user understanding the outer stimulus and using rules

[5] C.D. Wickens, S.E. Gordon, and Y. Liu, Intro. Human Factors Engineering, Longman, 1988.

Table 3. Degrees of User's Expertness and User Observation Data

Degree of Expertness	Behavior	User Observation Data		
		Perception	Recognition	Behavior
Novice	Analytical Behavior			
Intermediate	Intuitive Behavior	○	○	○
Expert	Automatic Behavior	○		○

and habit in the recognition process, the observed data are smaller. Especially, since the experts skip the analysis of stimulus from outside, and since they go directly into the task execution, the data collected from the recognition process stage would be insufficient.

In the integrated Perception/Recognition/Behavior evaluation system, the usage problems that vary on the degree of expertness can be classified as follows:

1. Novice users: usage problems can be found on analytic recognition stage, such as limited short-term memory, biases on developing hypothesis or action, cognitive clinger.
2. Intermediate level users: usage problems that are related to the mis-application of rules, based on misled decision making.
3. Expert users: Usage problems can be observed in the cases that users pay attention to unimportant places. On the contrary, sometimes users pay too much attention to the task itself. The over-concentration on the task disturbs the processes of the task.

4 EMT Tool: Integrated Evaluation Tool of Perception/ Recognition/Behavior

4.1 Development of EMT Tool

The EMT System collects the user's experience data and web environment data from the usability test experiment. The collected data is not analyzed individually, but instead it is connected to other kinds of data and the user's task progress situation as well as the recognition process of the user is analyzed. In order for this type of integrated analysis, the EMT System requires the EMT Tool; Integrated evaluation tool of Perception/Recognition/Behavior that can effectively manage and analyze various kinds of data.

The requirement for EMT Tool can be summarized into the following categories of data logging, data synchronization, visualization of generalized data, reduction of video analysis time, and memory recollection material.

4.2 Scenario of EMT Tool's Application

The EMT Tool is utilized from 'Step 3. Preparation' to 'Step 6 Results analysis' of the progress scenario of EMT System (refer to Fig. 2). In order to see the role and progress process of EMT Tool in a scenario, refer to Fig. 3. After the calibration, EMT Tool collects the user's eye and mouse tracking data and web environment data in the Test stage and stores the data. And in the Interview stage all the data are

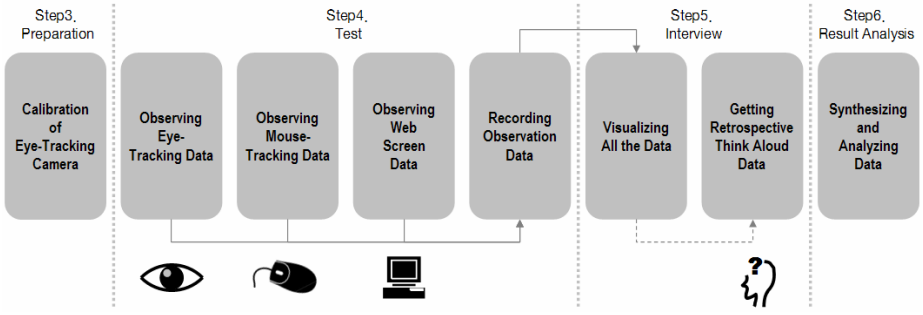


Fig. 3. Scenario of the EMT Tool

collectively reproduced visually while the user records the retrospective think aloud accounts using it as a reference.

4.3 Realization of EMT Tool

The EMT Tool is a tool that realizes the execution of the EMT System. The EMT Tool acts as the "Observance of the user's experience and web environment" and "the analysis of collectively observed data". To carry out these two tasks, EMT Tool consists of two different tools which are EMT Tracker, which is the observe/collect tool, and EMT Analyzer, which is the synthesize/analyze tool. The EMT Tracker is executed in the 1) zero point adjusting of the eye tracking device ~3) data storage stages of the tool scenario [Fig. 3]. The EMT Tracker's process is as follows: First the zero point is adjusted between the testee and the sight tracking equipment, and then it observes the user and the web to collect data, and finally stores this data in the data storage stage.

The EMT Analyzer is executed in the 4) Visualization of data ~ 6) Integrated Analysis stage [Fig. 3]. The EMT Analyzer loads the data that was stored by the EMT Tracker, carries out the analysis, the sight, mouse, and web screen data is collectively visualized, and the testee's Think Aloud description is inputted to provide an analysis frame.

The screen of the EMT Tracker is shown in [Fig.4]. The calibration and the starting point and the end button of eye-tracking recording are provided as a toolbar. The screen of the EMT Analyzer is shown in [Fig. 5]. In the background, the web screen video data is played while the user's eye movement path (red) and mouse movement path (blue) is shown above the screen video. In addition, 'EMT information palette' for data adjustment and confirmation, 'Time slider palette' for data play search, 'EMT analyzer palette' for consciousness/thinking/movement integrated analysis are provided. The user can move the palette and also can set it to disappear from sight and can also move the palette while the player is being played. Also, when a singularity is found, contents can be inserted at any time and place, and the place of insertion is indicated on the player bar so that an effective analysis is possible.

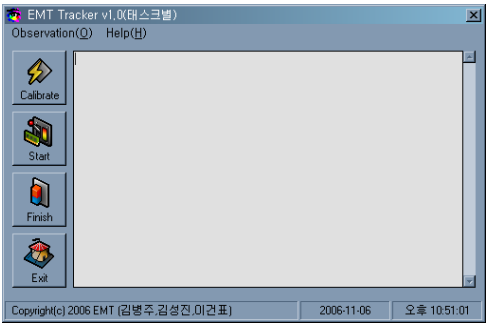


Fig. 4. EMT Tracker

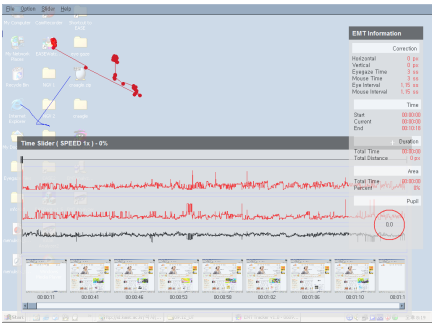


Fig. 5. EMT Analyzer

As it was described above, the EMT Tracker and EMT Analyzer of the developed EMT Tool send and receive data and progress linked to each other. [Fig.6]

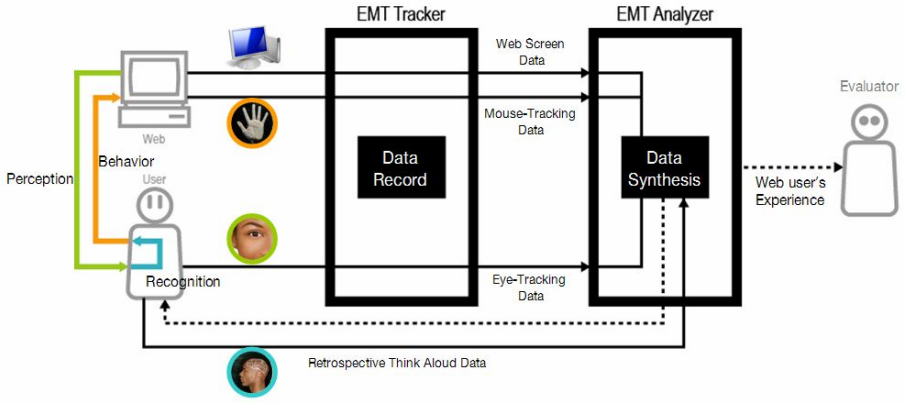


Fig. 6. Overview of EMT Tool

5 Conclusion and Future Study

For a new web usability evaluation method, this research proposes EMT System that can directly observe the web user's experience, and can analyze the user's potential thinking procedure, and at the same time EMT Tool was developed in order to process the evaluation system. If we elaborate in more detail, first, the user's experience is divided into perception, recognition, and behavior and each item's experience observation methods are as follows: eye-tracking for perception, retrospective think aloud for recognition, mouse-tracking for behavior. Second, EMT System was proposed from the generalized user's observation method for each detailed experience, and that process of usability evaluation experiment and data observation and the process of collection, and data analysis frame work were explained. Third, EMT Tool was developed to execute EMT System.

The EMT Tool, in comparison to other eye-tracking and mouse-tracking methods utilized in previous web usability evaluation methods, has the advantage where both eye-tracking data and mouse-tracking data can be shown on the web usage screen and that it is able to complementarily analyze two different data simultaneously. Also, it can spontaneously input retrospective think aloud accounts from revived memory of what the user saw and reacted during the process of web experience through eye-tracking and mouse-tracking data. EMT Tool is not a one-way interface where the evaluator questions the user. It is a bi-directional analysis interface where the user spontaneously records what he or she felt and thought during the web experience process and the evaluator analyzes it.

For future research, first, we require an improvement of interface for increase in usability of EMT Analyzer as well as an addition of result output function for increase in practical usage. Although it is possible to use various functions through a menu, since accessibility can be an issue when using a menu, a tool bar is needed for user's convenience. Also, for an immediate comprehension for the user, the text-centered interface needs to be modified to the icon-centered type. In addition, documentation of analysis results or a function to output contents in commercial file types for editing is required. Second, in order to verify the EMT System and the practical usage of EMT Tool and its wide usefulness, research on various user and environment in different cases is required. Even though EMT System and EMT Tool were developed for web usability evaluation purposes, it can observe web in computer environments as well as in various software. Therefore, research utilizing EMT System and EMT Tool in various situations must be carried out.

References

1. Park, C.M.: A Study on the Visulization & Analysis of User Interaction on the WWW, Master Thesis at KAIST (2001)
2. Wickens, C.D., Gordon, S.E., Liu, Y.: Intro. Human Factors Engineering. Longman, Longman (1988)
3. Donald, A.N.: The Design of Everyday Things, Doubled Currency (1988)
4. Gordon, S.E.: An Information-Processing Model of Naturalistic Decision Making. Presentation at Annual Meeting of the Idaho Psychological Association, Sun Valley, Idaho (1997)
5. Nielsen, J.: Designing Web Usability, New Riders Publishing (2000)
6. Trumbo, J.: The Spatial Environment in Multimedia Design: Physical, Conceptual, Perceptual, and Behavioral Aspects of Design Space, Design Issues, vol. 13(3)
7. Preece, J.: Human Computer Interaction. Addison-Wesley, London (1994)
8. Kantowitzm, B.H.: The role of human information processing models in system development. In: Proceedings of the human Factors Society 33rd annual meeting, Santa Monica, CA, Human Factors Society (1989)
9. Guan, Z., Lee, S., Cuddihy, E., Ramey, J.: The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: Proceeding of CHI2006 (2006)

Cultural Difference and Its Effects on User Research Methodologies

Jungjoo Lee, Thu-Trang Tran, and Kun-Pyo Lee

Human-Centered Interaction Design Lab, Department of Industrial Design,
Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu,
Daejeon, 305-701, Republic of Korea
givelo@gmail.com, {trangtt,kplee}@kaist.ac.kr

Abstract. Various researches have proved that cultural differences affect the process and results of user research, emphasizing that should cultural attention be given in order to obtain sufficient results. After performing three experiment methods: probe, usability test, and focus group interview in the Netherlands and Korea, we discovered that productivity and effectiveness was poorer in Korea. The differences were found due to the contrary between cultures, strongly indicated by Hofstede's cultural dimension Individualism vs. Collectivism. In addition, we have proved that the different factors made an impact on user research process and result. Based on the analysis, we compiled guidelines for each of the method when performing in Korea.

Keywords: Cultural difference, User research methodology, User participatory study, User research guidelines.

1 Introduction

As market competition globalizes, understanding users of various cultures during the design process has become important for securing a corporation's competitive power. However, most user research methods currently in use have been developed for people living in the US and Europe. It makes us wonder if those methods can achieve the expected sufficient results when applied to people living in other cultures.

Recently, various researches have proved that cultural differences affect the process and results of user research, emphasizing that should attention be given to users' various cultural backgrounds in order to obtain sufficient results (Hall, 2004). Yet there lacks of insightful studies concerning the relationship between culture and user research methods. This paper aims to reveal the effects that cultural differences have on user research and suggests guidelines as how cultural difference can be considered to conduct effective user research and gain useful results.

2 Study of Background Theory – Cultural Difference

2.1 Cultural Variables

Edward Hall's context theories say that information during a communication or information in a message is part of context (Hall, 1989). In high context culture, most

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 122–129, 2007.
© Springer-Verlag Berlin Heidelberg 2007

information is included in the context, thus it is less externally expressed. However, communication is direct, clear, and expressed externally in low context culture. Concerning Hall's context communication, Hofstede revealed that high context communication occurs in collectivistic culture and low context communication occurs in individualistic culture, according to his cultural difference's dimension of individualism vs. collectivism (Hofstede, 2001).

2.2 Culture and Politeness Theory

Based on Politeness Theory,[1] Ting-Toomey developed Facework Framework, explaining the difference in communication pattern in individualistic-low context culture (LCC) in which one desires not to be disrupted, intruded, and forced by others., and collectivistic-high context culture (HCC) in which one desires to be liked and approved by other people and concern about others' reaction.

3 Relationship Between Cultural Difference and User Research Methods

3.1 Classification of User Research Methods Regarding Communication Pattern

It can be said that user research methods collect knowledge in different levels according to the characteristic of the communication. This knowledge about user experience can be distinguished by possibility of observation and explicitness. Sanders (1999) explained that in order to effectively observe knowledge at different levels, different methods must be applied according to the characteristics of that knowledge level, as shown in figure 1.

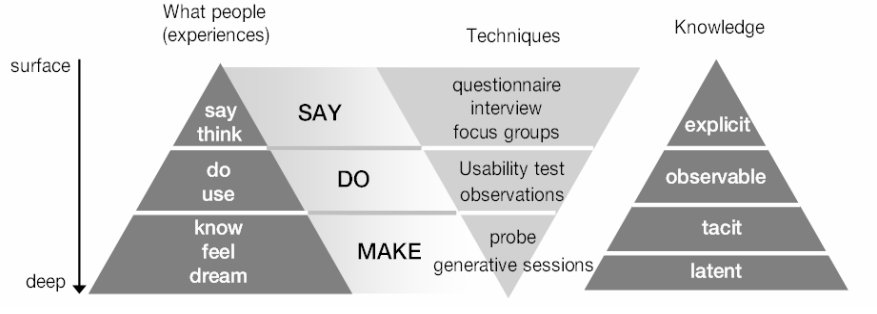


Fig. 1. Cognitive level of user experience and corresponding methods of user research with different communication patterns

3.2 Extraction of Influential Factors

The characteristics in different cultures were integrated and mapped to communication patterns of user research. Group activity was also mapped together to

[1] Politeness Theory explains that social intimacy, politeness, or implicative expression is emphasized when a speaker demands a listener in general (Brown et al, 1987).



Fig. 2. Integration of cultural difference and extraction of influential factors

take into account some cases where the group constituent was one of the targets of face-keeping.

4 Results and Analysis

4.1 Experiment Design

Participant Selection. According to Hofstede's cultural dimensions (Hofstede, 1991), the Netherlands (individualism figure of 80) was selected to represent individualistic-LCC, and Korea (individualism figure of 18) to represent collectivistic-HCC. Six engineering major university students from each country, who are in their 20s were selected, male:female ratio was 3:3, and none of the participants had previous experience with any of the tests.

Experiment Methods. In this research, probe, usability test, and focus group interview were conducted to observe the effects from four factors mentioned in 'Extraction of Influential Factors'. 'Design of next generation's portable media device' was selected as the topic of the experiment, considering the diverse perceptions in each country.

Probe - to observe how participants act to ambiguous and open-ended tasks during workbook writing and photographing, we provided very expandable and self-interpretable tasks that can highly reflect an individual's own experience. Participants are encouraged to escape from formality of writing so they can make use of the free-form (with aid from photograph, drawing, stickers, etc.).

Usability test - to observe participants' eagerness to find problems as well as verbal protocol and attitude during the usability test, participant was allowed to talk about the product's problems (Iriver U10 and Sony PSP) after using them for seven various given tasks.

Focus group interview - find out how comfortable a participant when sharing his own experiences and thoughts in a group. In addition, it was used to observe how easily a participant can be affected by the opinion of the majority and interaction among group members.

4.2 Results

User research was performed firstly in the Netherlands at Delft University of Technology and then in Korea at Korea Advanced Institute of Science and Technology.

Probe. Participant's feedback during probe period - the procedural aspect, and sufficiency of workbook writing and photography - the result aspect, was analyzed.

Participant feedback. Even though both groups felt the ambiguity of terms on the workbook, they attempted to interpret those ambiguous terms own their own to complete the task without any help. Dutch participants wrote in the workbook almost every day, but Korean participants later revealed that they had trouble writing in the workbook every day, so sometimes they wrote several days of work all at once.

Sufficiency. Workbook contents and photographs were analyzed and compared to find out how sufficient each group expressed their experiences and how diligent they were in taking photographs. Dutch participants' sufficiency was found higher than Korean participants'. Not only Korean participants gave only short answers to workbook questions, but were also poorer in exploring the freeform expression.

Usability Test. Protocol analysis was performed on interview and behavior of participants to compare tendency to criticize a problem and attitude towards participation.

Eagerness of Usability Test. Figure 4 shows that Dutch participants criticized the products more actively, seeing a product's both weaknesses and strengths.

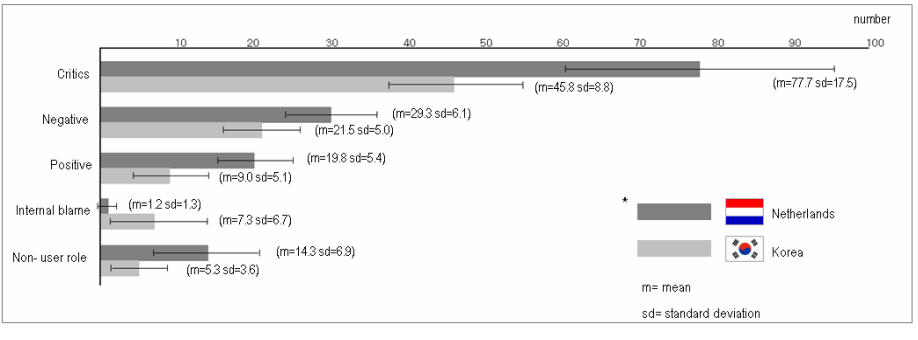


Fig. 3. Comparing performance during Usability test between two groups

Tendency towards Self-Criticism. Dutch participants believed that most problems that occurred during the tests were product faults. Relatively speaking, Korean participants blamed their mistakes for the problems occurred. However, it varied greatly from individual to individual (Mean =7.3, SD = 6.7), discrediting the conclusion that Korean group have more tendency towards self-criticism. Presumably, the participants were well-educated engineering students thus they were comfortable with the whole test situation and handling digital products.

Level of Anxiety during Usability Test. In general the level of anxiety did not differ much between two groups of participants.

Diligence of User Role. Korean participants maintained user role better than Dutch participants. Dutch participants explored product functions outside given tasks and sometimes criticized the tasks. Miranda Hall's research has shown that Dutch participants had a wider range of observation and also discovered a wider range of problems, not to mention their frequent escape from the user role.

Focus Group Interview. Protocol analysis was done according to the timeline to gather for all participants' frequency of presenting an opinion, interaction style, and to observe the role required by the moderator.

Active Participation and Even Distribution of Voice. Figure 5 conveys that Dutch participants engaged more actively in the discussion, continuous expression of thoughts and ideas without much help from the moderator. The Korean timeline [Fig.6] shows that there are temporal spaces between opinions and seems like responses were given only when the moderator asked a questioned or pointed out someone to speak. Since Dutch participants were more active when suggesting opinions, they were also more likely to escape from their user role in comparison to Korean participants. Korean participants were found to speak more frequently as time elapsed and that they need a refreshing time to increase the rate of participation.

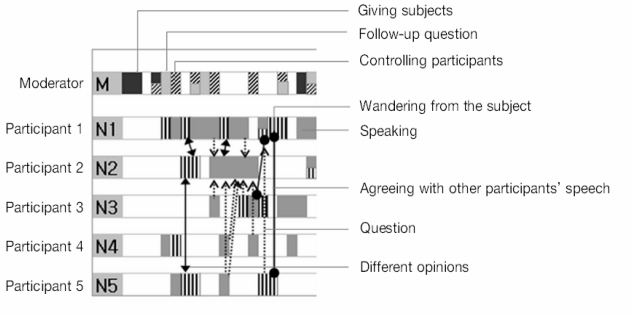


Fig. 4. Description of timeline diagram

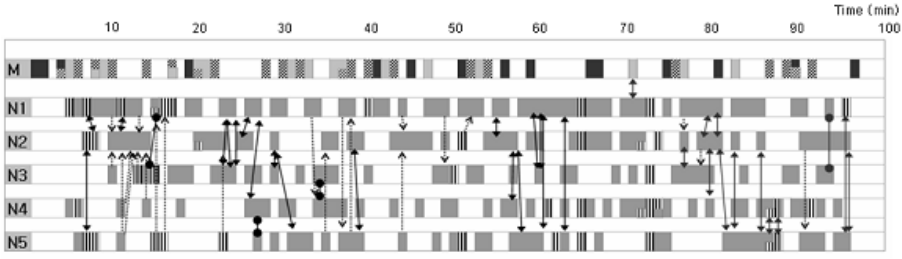


Fig. 5. Analysis of Dutch participants' focus group interview timeline

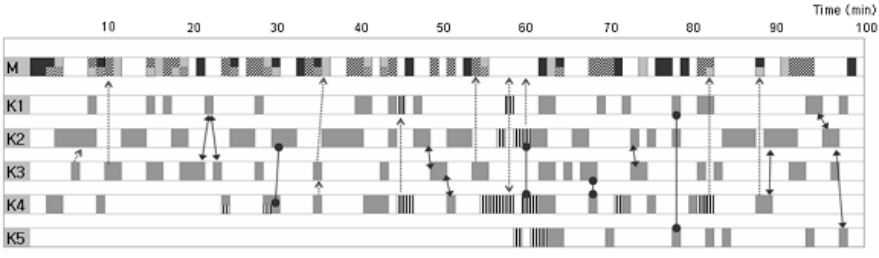


Fig. 6. Analysis of Korean participants' focus group interview timeline

Role of the Moderator. In Korean group, moderator needed to specifically assign on participants constantly and ask more detailed questions to carry on the discussion. On the other hand, Dutch moderator did not have to do much since participants actively engaged in discussion as soon as the discussion topic was suggested. But some participants tended to become verbiage or circulated the same topic for too long, requiring the moderator to control such behavior.

Interaction among Participants. Voluntary interaction amongst group members was more obvious in the Dutch group, when someone finished, opposing opinion and corresponding questions were actively generated. On the other hand, Korean participants were apt to ask the questions to the moderator. Assumingly the high uncertainty avoidance causes such behavior; constituents are less likely to engage in free discussion but are more likely to seek for confirmation from the moderator. There was no significant conformity of opinion in either group.

4.3 Discussion

After performing three experiment methods in the Netherlands and in Korea, we discovered that productivity and effectiveness was poorer in Korea. The differences lay in spontaneous participation, uncertainty avoidance, style of problem criticism, and attitude were found due to the contrary between Individualism-LCC and Collectivism-HCC. In addition, we have proved that these factors made an impact on

user research process and result. Based on the result, we have compiled guidelines for each of the method when performing user research in Korea.

Probe. Constant communication between a participant and a researcher is crucial during the probe period to boost participant's motivation and stimulate of responsibility. The communication between a researcher and a participant must be playful and informal. To alleviate any burden from the participant and induce fun, some playful tasks and factors must be added to the probe tool. Besides, the design must be well designed and some "cute" and "friendly" factors should be augmented so that participants can feel more comfortable and friendly.

In order to increase motivation, should task planning be more considerate of a specific cultural zone by finding out local activities in trend or spread of technology.

Usability Test. To increase the efficiency, some type of orientation or sensitizing process must be provided to teach participants to have a critical mind. Attempt to switch to less direct method to find problems rather than face-to-face interview.

Focus Group Interview. Game factors such as warming-up sessions before the interview and a game to increase friendliness among participants are needed. Provide devices that will make the participants feel obligated to speak (ex. microphone) or come up with factors that will promote detailed explanation of one's opinion.

5 Conclusion and Future Research

This paper has analyzed how cultural difference affects the user research process and result and suggested guidelines how user research in Korea should consider cultural effects.

Nonetheless, the limitation of this qualitative research lies in that the sample was extremely small. Besides, the participants do not sample the general population since they were student in their 20s from highly educated engineering schools. Therefore, this paper can become the foundation of future research, which will aim to include a wide range of age groups and numerous participants. If this research continues on, valid data of various cultures will become available. Moreover, the guidelines of considering cultural effect in user research suggested by this paper will have to prove its usefulness by cases of real life applications.

References

1. Brown, P., Levinson, S.: Politeness: Some Universals in Language Usage. Cambridge University Press, Cambridge (1987)
2. Hall, E.: Beyond Culture. Anchor Books Editions (1989)
3. Hall, M., et al.: Cultural Differences and Usability Evaluation: Individualistic and Collectivistic Participants Compared. Technical Communication. vol. 51 (2004)
4. Hofstede, G.: Culture and Organizations: Software of the Mind. McGraw-Hill International. The United Kingdom (1991)

5. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nation. 2nd edn. Beverly Hills (2001)
6. Rijn, H., et al.: Three Factors for Contextmapping in East Asia: Trust, Control and Nunchi (2006)
7. Sanders, N.: Design for Experiencing: New Tools. In: Proceeding of the First International Conference on Design and Emotion (1999)
8. Ting-Toomey, S.: Intercultural Conflicts Styles. A Face-negotiation Theory. In: Kim, Y.Y., Gudykunst, W.B. (eds.) Theories in Intercultural Communication, Sage, Thousand Oaks (1998)

A Development of Graphical Interface for Decision Making Process Including Real-Time Consistency Evaluation

Joong-Ho Lee, Ki-Won Yeom, and Ji-Hyung Park*

Intelligence Interaction Research Center, Korea Institute of Science and Technology, HaWolGok 39-1, Seoul, 130-650, Korea
jhpark@kist.re.kr

Abstract. Decision making problems are often imprecise and changeable because of potential inconsistency in human thinking. Although AHP gives a desirable guide to the reasonable solution via consistency ratio, there is still possibility of containing inconsistence during process. Therefore, an important step in many applications of decision making problems is to perform a consistency analysis in real-time. We introduce a new method of priority setting in decision making processes, which is implemented as an interactive and convenient graphical interface of the decision making problem. It is designed to support the real-time consistency evaluation. The conventional AHP does not provide graphical user interface and is impossible to monitor the interim findings in the middle of process, and is difficult to predict the difference of results when changing pair-wise comparison conditions, and is difficult to monitor the consistency of human judgment during operation. The proposed real-time calculation algorithm and visualization method is developed to realize effective and reliable decision making environment, and is verified its merit through the exemplary case. In addition, we propose new algorithm of evaluating consistency level. The rationality tension is proposed as a new index for evaluating a real-time consistency analysis with interactive graphical user interface. It is desirable for a system to provide fast and visible information of consistency in decision making processes.

Keywords: Decision Making, Priority Setting, AHP, Visualization, Interactive process, Consistency Ratio.

1 Introduction

The analytic hierarchy process, AHP [1] is a comprehensive, logical and structured framework for solving a priority setting. It improves understanding of complex decisions by decomposing the problem in a hierarchical structure. This method incorporates all relevant decision criteria and allows the decision maker to determine the trade-offs among alternatives by pair-wise comparison.

* Corresponding author.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 130–137, 2007.
© Springer-Verlag Berlin Heidelberg 2007

1.1 Decomposition of the Decision Problem

At first, User defines the problem. Fig 1 shows the exemplary hierarchy of the problem. To acquire the best choice among alternatives, user should define decision criterions. In other words, the problem is which alternative cloud be the best choice to meet the goal considering all criterions.

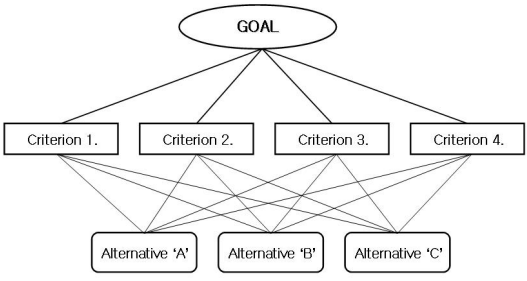


Fig. 1. To find the best choice for a desirable GOAL, user defines criterions and derives priority setting considering each criterion. The above shows exemplary decomposition of four criterions and three alternatives.

1.2 Pair-Wise Comparison Between Alternatives

AHP gives an effective evaluation methodology – pair-wise comparison, which guides users to decide priorities of alternatives by considering only two alternatives at one time. Fig. 2 shows exemplary AHP decision table. All pair-wise comparison values are filled up on the table. This table is called comparison matrix (M). Usually, it is recommended to use 1, 3, 5, 7, 9, 1/3, 1/5, 1/7, 1/9 for the value.

* Criterion 1's comparison table.

	Alternative 'A'	Alternative 'B'	Alternative 'C'
Alternative 'A'	1	3	1/5
Alternative 'B'	1/3	1	1/7
Alternative 'C'	5	7	1

$$M = \begin{pmatrix} 1 & 3 & 1/5 \\ 1/3 & 1 & 1/7 \\ 5 & 7 & 1 \end{pmatrix}$$

Fig. 2. The pair-wise comparisons are filled up on the table. This table is called comparison matrix M.

1.3 Synthesis of the Priorities

Synthesizing the comparisons is to get the priority setting of the alternatives with respect to each criterion and the weights of each criterion with respect to the goal. User calculates eigen values or geometric averages for each alternatives. Local priorities are then multiplied by the weights of the respective criterion. Finally, the results are summed up to get the overall priority of each alternative. The local priority setting of the above table in Fig. 2 is shown on table 1.

Table 1. The local priority setting of criteron 1

	Alternative 'A'	Alternative 'B'	Alternative 'C'
Weight	1.182	2.759	0.306
percentage	27.8%	65%	7.2%
Rank	2	1	3

2 Consistency Evaluation in AHP

The consistency index C.I is used for evaluating rationality of the AHP priority setting result. If C.I is too large the result is regarded to be unreasonable because there may be too much antinomy in comparison. C.I is an important index to verify the consistence of decision making process.

2.1 Consistency Index

As a result above in Fig. 2, Alternative 'B' is the best decision(65% of priority). Since 'B'>'A' and 'A'>'C', logically, we hope that 'B'>'C'. This logic of preference is called transitive property. If user's judgment of comparison 'B' vs 'C' is transitive, then this decision process can be considered as consistent one. On the contrary, if the user judged 'C' is more important than 'B', this decision has inconsistent judgment. To evaluate the decision's consistency, a numerical index is used, which is closely related to the transitive property. A comparison matrix M(as shown on Fig. 2) is said to be consistent if,

$$\mathbf{m_{ij} \times m_{jk} = m_{ik}} \quad \text{for all i, j, k} \tag{1}$$

For consistent reciprocal matrix, the largest eigen value is equal to the number of alternatives, or $\lambda_{max} = n$. Then a measure of consistency is below,

$$\mathbf{CI = (\lambda_{max} - n)/ (n - 1)} \quad \text{where } \lambda \text{ : eigen value of matrix M} \tag{2}$$

n : number of alternatives

Thus in our previous example shown on Fig. 2, we have λ_{max} =3.054 and three comparisons, or n=3, thus the consistency index is 0.027.

2.2 Consistency Ratio

The consistency level is measured by comparing it with the appropriate one. The appropriate consistency index is called random consistency index (R.I). It is randomly generated reciprocal matrix using scale 1/9, 1/8, …1, … 8, 9. The random consistency index is shown on the Table. 2.

Table 2. The random consistency index R.I with respect to size of matrix M

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

The consistency ratio(C.R) is a comparison between consistency index(C.I) and random consistency index(R.I), or in formula,

$$\mathbf{C.R = C.I / R.I} \tag{3}$$

If the value of consistency ratio is small enough or equal to 10%, the inconsistency is acceptable. If the consistency ratio is greater than 10%, we need to revise the subjective judgment. For our previous example, we have C.I=0.027 and R.I for n=3 is 0.58, then we have, 4.66% for C.R. Thus, this subjective judgment can be considered as consistent process.

3 Real-Time Priority Setting

The real-time priority setting that we introduce in this paper is a new method of decision making problem. It is composed of a real-time calculation algorithm and visualization. As the computer environment growing fast, a study for interactive operation between human and computer for decision making problem is newly issued. The real-time priority setting is suitable for up-to-date computing environment to complement defects of AHP. The real-time priority setting algorithm replaces stepwise manual calculation process of AHP. This algorithm can refresh interim finding immediately after the change of pair-wise comparisons. An interim finding is shown on display with graphical symbols which increase visual understanding. User can sense current situations directly via graphical expressions. And when the conditions changed, for example, new emergence of additional alternative, user can confirm the updated result only by handling a few graphical factors instead of remaking new AHP table and restarting the process from the beginning.

3.1 Graphical Expression

A conventional AHP shows the result only by numerical table, which makes user hard to understand the result instantly. It is certain that more legible information with graphical expression improve the efficiency significantly. The Real-Time Graphical Expression gives not only numerical information but also symbolic expression such as circle, line, color, size and thickness. It gives information of current status for users to understand interim findings rapidly and easily. Fig. 3 shows the exemplary symbolic expression of the real-time priority setting.

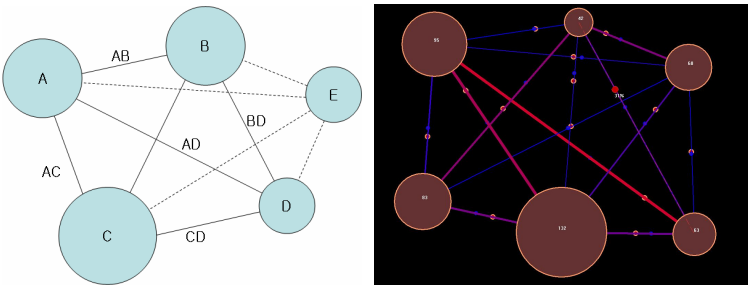


Fig. 3. The real-time priority setting method uses graphical expression as shown. Using circles, lines and their sizes, colors, the interim finding can be monitored interactively.

Circles represent alternatives. They have basic information of alternative's features, remark, keyword, weight percentage. Each size of the circles represents its current weight. As a pair-wise comparison judged, a line is created by human input gesture. The line represents comparison value and its consistency level. During decision, users may create many notes. The system allows users to edit remarks on the circle.

3.2 Algorithm

The real-time priority setting method derives the weight of alternatives via iterative linear algebra. This method uses the algorithm which result is proved to be nearly corresponds to conventional AHP method. A case shows prioritizing among 3 alternatives (A, B, C). When a pair-wise comparison is decided, an interim finding is updated immediately as below.

$$P^n_A = (P^{n-1}_A + R_{AB} \times (P^{n-1}_B + P^{n-1}_A) + R_{AC} \times (P^{n-1}_C + P^{n-1}_A)) / N \quad (4)$$

Where, P^n_A is the weight of alternative 'A' in n's iteration, R_{AB} is the priority value between 'A' and 'B'. N is the number of terms of algebra. As the iteration repeated (if n is sufficiently large), a difference between P^n_{all} and P^{n-1}_{all} comes to decrease and P^n_{all} converges to the unique value.

$$P^n_{all} - P^{n-1}_{all} < Z \quad (5)$$

Iteration halt constant Z reduces the number of iteration. If Z is sufficiently small, final P^n_{all} corresponds to the unique value. Table 2 shows the result of real-time priority setting for the same comparison matrix M in Fig. 2. Furthermore since this calculation can be done even though not all pair-wise comparisons decided, user can calculate interim findings. For example, if R_{AC} is not decided yet, it is set to zero. The result could express reasonable interim finding while the conventional AHP can not show any reasonable index in the middle of the process.

4 Rationality Tension

The consistence index C.I is used for evaluating rationality of the AHP result. If C.I is too large the result is regarded to be unreasonable because there may be too much antinomy in comparison decision. In this paper, new index called Rationality Tension (R.T) is introduced for evaluating the consistency of the result.

4.1 Algorithm

Since a human thinking has a possibility of containing inconsistency along to the decision making, rapid interaction for informing that he/she is going with too much inconsistency could provide considerable advantage. When the pair-wise comparison is changed, K value is updated as below.

$$K_{AB} = | R_{AB} - P^n_A / P^n_B | \quad (6)$$

K_{AB} is an index of difference between current pair-wise comparison between 'A' and 'B'. On condition that iteration of real-time calculation is enough to acquire the

valid value, a priority is stable on balance of itself and other's effects. If all K values of alternative 'A' are small enough for all other's relations, there are small antinomies with respect to the alternative 'A'. While AHP considers only one representative value for a matrix M, this algorithm makes it possible to consider all relation's consistency levels case by case.

$$R.T = \text{Max}\{K_{AB}, K_{AC}, K_{BC}, \ldots\} \tag{7}$$

R.T is 'rationality tension' which is proposed in this paper for a new index of consistency level. The rationality tension R.T corresponds to C.I. It can be calculated at any stage of process. User can monitor the R.T to check the validity of current decision whenever he/she needs to inspect whether the relation keeps too much contradictions or not. A Monitoring of R.T is available on display with various color and thickness of line. As the R.T increases large, color and thickness are changed to be detected clearly by user. User can modify the pair-wise comparison judgment which includes high R.T. This advantage leads users to qualify their conclusion more effectively.

4.2 Visualization

If a comparison line has low R.T, it is expressed as the thin and dark colored. On the other hand, a comparison line which has high R.T is expressed thick with deep red color as shown on Fig. 4.

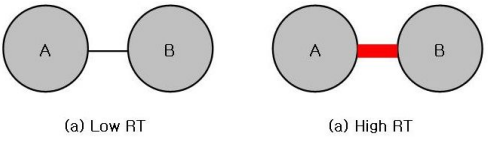


Fig. 4. The Real-Time Graphical Expression shows the current consistency information R.T with thickness and color of line for user to understand instantly.

When user operates decision making with pair-wise comparison, the real-time priority setting gives information about R.T instantly. It would help users to find where the inconsistency occurred significantly during decision. And it leads users to modify comparisons to the reasonable consistency level.

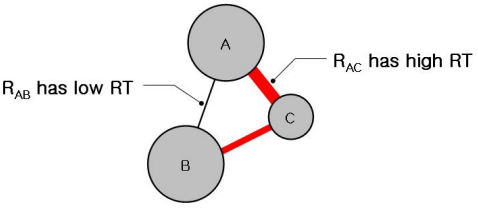


Fig. 5. Real-time expression of consistency levels helps user to derive reasonable result in decision making

4.3 C.R(Consistency Ratio) vs R.T(Rationality Tension)

According to our study, C.R= 10% corresponds to R.T=10 approximately. If R.T is increased up to 10, the depth of line becomes thick and color becomes red, which means that user should consider previous comparisons again. Fig 6 and Table 3 shows the results done by real-time priority setting of the matrix M in Fig. 2.

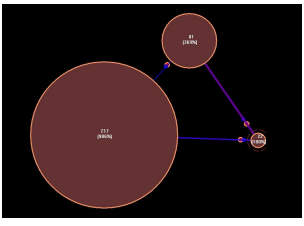
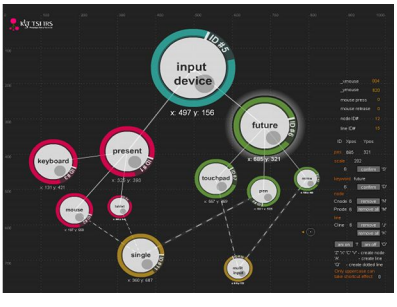


Fig. 6. The real-time priority setting result of the matrix M in Fig 2

Table 3. The real-time priority setting shows approximately same result to that of AHP

P_A	P_B	P_C	K_{AB}	K_{AC}	K_{BC}	R.T
25.3%	65%	7.2%	2.2	4.4	2.8	4.4



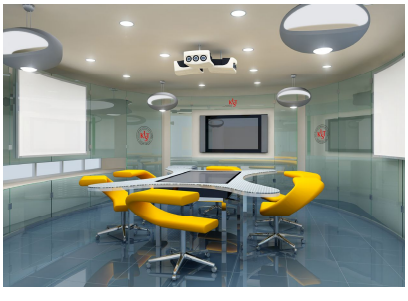


Fig. 7. The Graphical Interface for Decision Making System has installed in IRS as an interactive solution for supporting the priority setting

5 Conclusion

The real-time priority setting enables users to monitor interim findings with interactive user interface. Symbolic expression improves visual understandings. And when defining pair-wise conditions, user is able to evaluate whether the decision retain the desirable consistency. It is more effective and fast than conventional AHP process. Finally this method proved to be suitable for remote, collaborative, large displayed environment by sharing visual symbolic information among users. And it provides users with real-time responses relative to users operation. This method has been substantially adopted into IRS(Intelligent Responsive Space) digital workbench

which is being developed as a Tangible Space Initiative project at Intelligence Interaction Research Center in KIST(Korea Institute of Science and Technology).

References

1. Saaty, T.L.: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, pp. 56–57. McGraw-Hill, NewYork (1980)
2. Saaty, T.L.: Decision Making, Scaling, and Number Crunching. Decision Sciences, pp. 404–409 (1989)
3. Park, J.-H., Lee, J.-H., Yeom, K.-W., Lee, S.-S., Eom, J.-I.: Graphical Expression Method for Decision Process Support, HCI2006 (2006)
4. James, S.D.: Remarks on the Analytical Hierarchy Process, Management Science, pp. 249–258 (1990)
5. Saaty, T.L., Jo, G.-T., Hong, S.-W., Gouen, C.-S.: Decision Making for Leaders, RWS pub. (1995)
6. Escobar, M.T., aguaron, J., Jimenez, J.M.M.: A note on AHP group consistency for the row geometric mean priorizatio, European Journal of Operational Research
7. Escobar, M.T., aguaron, J., Jimenez, J.M.M.: Consistency stability intervals for a judgment in AHP decision support, European Journal of Operational Research

Using Webzine to Create Effective Communications Between China and the West

Christina Li, Sean Liu, and Eleanor Lisney

uiGarden Project Team, Infinite Interactive Ltd. UK
{Christina.li,sean.liu}@uigarden.net
elisney@gmail.com
http://www.uigarden.net

Abstract. Knowing the development and opinion from other nations is essential for designing usable product for different cultures. Effective communications between different countries is invaluable, however, often inhabited by the problem of limited language access. This paper will provide insight and practical experiences about how we offer swift and free information exchanges between UI practitioners in the west and in China by a bilingual webzine, - uiGarden, that provides an opportunity for researchers and practitioners who work in the user experience design field in the Chinese and the English speaking worlds to exchange views and deepen each other's knowledge in the field.

Keywords: cultural exchange, China, cross-culture usability, webzine.

1 Introduction

With an ancient civilization and a population of 1.4 billion people, China has become one of the fastest-growing countries in the world. In the economic arena, it has a rapid growth of 9.5% per year in the past two decades. China is becoming potentially the biggest market and an economic giant in the near future. Therefore, China has attracted attention and investments from all over the world. With the rapid growth of the Chinese economy and the process of globalisation in recent years, Chinese enterprises realized that they had to strengthen their competitive edge to be able to survive and compete in the future.

At the same time, more and more multinational companies have entered the Chinese market. There is also a flow of people out of China. There are more and more Chinese students, businessmen, practitioners and researchers in immigrating countries like Canada, the United States, as well as Europe. Domestic enterprises in the west also have customers with Chinese backgrounds. These two factors have brought about a rapid increase in demand for usability.

The fact of encountering contracting cultures can be an enriching facet of modern life with intercultural exchanges, the proverbial 'melting pot' but also bewilderment, misunderstandings and miscommunication. Usability professionals both at the west and China are facing new challenges brought by the inosculation between the east and the west when designing usable and enjoyable user experience for their users.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 138–145, 2007.
© Springer-Verlag Berlin Heidelberg 2007

HCI and user experience design has captivated more and more designers, usability specialists, information architects, software engineers, and cognitive scientists, etc. to step into its field. There are 32 HCI consultancies registered on the British HCI group's web site [1] and 242 companies registered on HCI Bibliography web site [2]. Every year more than 50 conferences related to HCI are held around the world; about 100 books and 400 articles are published [3].

HCI and usability emerged in China much later than it did in the west, really only establishing itself as a field after 2000 and especially more significantly since 2003. Usability practice in China started from activities conducted by multinational companies, some setting up usability groups. Stiff international competition and the desire for development have also made user experience an important issue for many leading Chinese companies, some maintain usability groups of over twenty people and have integrated user-centered design (UCD) into their processes. Usability methodology in design is also being used in Chinese domestic large-scale enterprises. For example, focus groups and usability testing were used in the product design and develop cycle of the Software product, - 'Happy Home' of Lenovo group Ltd., the biggest IT Company in China.

However the growth of the usability field in China and a community of interest has led to the formation of professional organizations. Founded in 2004, ACM SIGCHI China [http://www.hci.org.cn], sponsors an annual national conference. UPA China [http://www.upachna.org] was set up in 2004 in Shanghai and organizes the User Friendly conference every year. The European Union-funded Sino-European Systems Usability Network project [http://www.sesun-usability.org] has organised four seminar and workshop tours around China and conducts joint usability studies in China. The first Harmonic Human Machine Environment conference (HHME) was held in October 2005; approximately 200 people, mainly from computer academia around China, attended.

2 Cultural Differences

Significant cultural differences exist within each country. China's population consists of 56 officially recognized nationalities, with the Han nationality (94%) most numerous. Although there are many different local dialects and accents, Chinese writing is uniform throughout the country, owing to the government's long-standing efforts to unify the language.

The global HCI community's understanding of the practical relevance of cultural issues in HCI has mirrored the timescale of the development of the subjects in China. In the last ten years HCI practitioners have changed their approach significantly to embrace cross-cultural development. When defining culture, researchers often refer to patterns of values, attitudes, and behaviours which are shared by two or more people. They further point out that culture is socially acquired, and that relationships with other people, relationships with the environment, and assumptions in term of space and language (for example) affect and shape culture, and are themselves affected by the culture [4]. Culture remains difficult to study, alone and certainly in relation to HCI practices. It is in particularly difficult to identify meanings, attitudes and expectations, not to mention the deeply embedded values and beliefs behind people's

thoughts, behaviours and actions. People's behaviours might be influenced by other factors (e.g., environmental conditions) rather than by their cultural traits, and the reasons for, and meaning of, an action can seldom be observed wholly and directly.

Studies in cross-cultural HCI have often embraced some consideration of cultural cognitive models. The model of national culture proposed by Hofstede has been frequently involved in the intercultural study of the use of systems [5]. However being based on a study of IBM employees in the 1970s there are significant gaps for the developing world.

Most Western software developers would support the principles of user-centred design but underlying concepts and assumptions are derived from USA / Northern European cultures. It is inevitable that those tools and techniques which involve users the most would be those very techniques which were most sensitive to cultural issues and the most susceptible to misinterpretations which could have serious impact on the quality of communication between designers and users.

Up to now, both non-native and native personnel have practiced usability in China [6]. The former were mainly involved in projects conducted by multinational companies, and their projects were usually supported by local recruiters, translators, moderators, and facilities. However, valuable information is sometimes hidden in subtle cues or deeply rooted in the social and cultural background, so barriers of language and culture can make a difference in usability studies. With the growth of local usability expertise, the "localization" of usability practice is an inevitable trend, and it will be reinforced by a difference in personnel costs.

3 uiGarden Webzine

3.1 Motivation

The geographical world is shrinking virtually everyday. People encounter different cultures as a matter of course in their daily lives. As user experience designers who have worked both in the east and in the west, we would like to help weave usability and cultures together. There's much potential to explore the Chinese market, - nearly one forth of the world's population [7]. On the other hand, we also want to bring the newest research and development in the west to China, in order to help the Chinese design and usability industry to keep up with the world.

The clamour is for rapid information access and exchange. Consumers are click happy and they expect instant feedback. Savvy web users, no matter where their geographical locations are, have no patience when it comes to net access. Even so, it is essential in designing for other cultures to understand their customs, habits and taboos. It is imperative to be aware of the particularities of the targeted community. There's also a demand for sharing experience with others. There is a plethora of on-line communities; discussion boards and mailing lists for exchanging ideas and thoughts. However, language barriers make it still difficult on exchanging information and experiences with the Far East. For example, at the Interaction Designers discussion group, many list members have complained that there's not enough research on other cultures and people make mistakes when developing product to the Far East market. Designers need to communicate with peers from other cultures.

Although the number of people in China dedicated full-time to usability practice is still small, maybe around 400, many product designers and developers are interested in usability. They are young, full of enthusiasm, and eager to learn. Of the people who are most interested in usability, quite a few are from design backgrounds with some others from programming background, probably because many companies employ design-trained people for user-interface design jobs.

At the first Sino-European Usability Seminar Tour held in Beijing, Shanghai, and Shenzhen, more than 200 people attended, with eighty percent from industry. Several companies sent more than ten of their employees to the event. A survey we conducted during the tour revealed that most of these companies have set up usability-related positions and departments. The respondents said they believed that usability would become more important in their organisations and that the major challenges at the moment are to master usability practices and skills and then to get their work recognized by their bosses and product-line units. Among all the aspects of knowledge sharing, Chinese practitioners are particularly long for customised practical instructions on methods of 'best practice' within the west for the development of on line applications (and indeed other areas) in the Chinese context that will not only make them be aware of UCD theory but also benefit their daily activities at work. Therefore, they wanted to attend training courses and learn from case studies so as to be able to start practicing usability in their daily work quickly.

However, in contrast to the rising expectation of theory and methodology practice, there are less than 20 research labs conducting HCI research in China and only a few of these is doing usability research and there is still no dedicated journal or conference in the subject of usability [8]. The limited existence of this area inevitably had some negative influence on the status of usability. There is also a gap between research and practice in the Chinese UI community. Except foreign enterprises, only a limited number of large-scale domestic Chinese companies have connection with researchers in the academic field. Most small to middle-scale IT or Dotcom companies don't take any research activity in their daily work. Therefore designers in those companies find difficult to improve their skills and to meet customers' needs. They have also claimed that lacking in specialised publications, swift trade information and lateral communication in the industry, they are finding it difficult moving ahead whilst trying to develop [9].

3.2 The Birth of uiGarden

When the Chinese UI design industries are increasing their understanding of usability concept, having found the lack and requirements of communication between the two community, with professional technology and experience in both the west and the east, a group of practitioners and researchers have started to dedicate themselves as a 'bridge' between the western and Chinese usability community.

uiGarden was conceived out of the collaborating team's passion for user-centred design. We recognise the need for a free flow of exchange between the Western and Eastern usability practitioners. We see the uiGarden to become an open platform, a potential for swift and free information exchanges between professional people who might not otherwise have the occasion to meet. There are undoubtedly cultural differences and, there are partly due to the language issues but not entirely so.



Fig. 1. The portal page to get into the site with different language

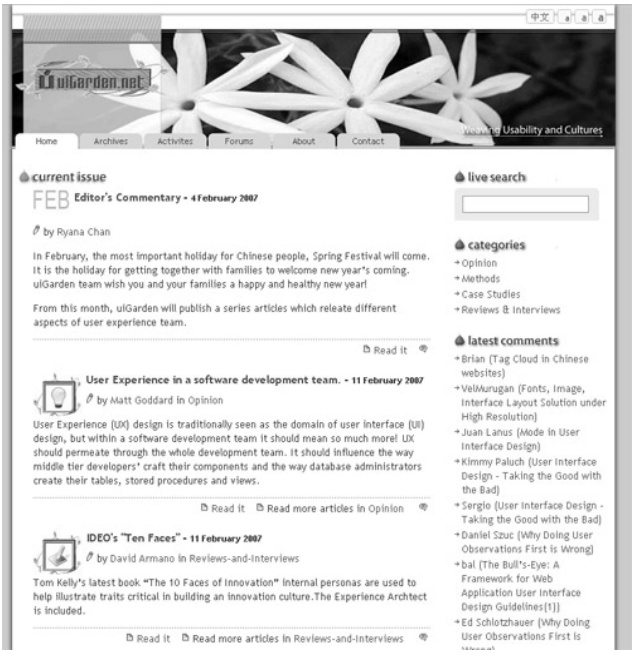


Fig. 2. The webzine in English

uiGarden is in the unique position of being able to discuss, explore and perhaps, resolve some of these differences. It can act as a catalyst to highlight the diversity of the websites between the West and the East.



Fig. 3. The webzine in Chinese

The team's goal is to provide swift and abundant information exchange and communication between the Chinese and western design communities, and to create a project that will mutually benefit both sides eventually. The birth of uiGarden means that communication between the west and the east has steeped into a new phase.

For western practitioners, we aim to create a place to communicate and exchange views with other professionals. The site also acts as a window on the user interface design industry in the Far East, helping to give insights into this increasingly important market.

From the point of view of Chinese practitioners, the website gives access to the latest developments in the West, featuring articles from leading Western experts translated into the Chinese language and providing discussion boards which facilitate discussion of each article.

Main contents are categorised by the type of them, - Opinion, Methods, Case studies and Reviews and Interviews. We bring audiences articles with a focus on:

- exploring theories and concepts that reflect current industrial practice
- future looking articles that address the challenges faced by our discipline
- articles relating to the teaching of user-centred techniques and methodologies
- case studies from projects demonstrating the application of user-centred techniques from two communities
- reviews of books, conferences, sites, software, tools and interactive projects

– interviews with leading experts in the field showing their point of view to professional issues

Articles in this journal are featured in two languages, English and Chinese. Discussion boards are provided at the end of each article to enable readers communicate with the author. Discussions are also translated to the other language in order to give readers from the other community opportunities to communicate with the author.

Besides the direct communication on the articles with their authors, we also provide forums in both language for casual discussion, focuses on various topics, not limited to one article. Popular topics are also being translated to another part, so participators will know what the popular issues are the most concerned by peers in the other community.



Fig. 4. Forums in two languages

This project has attracted attentions from all over the world and has got support from Britain's academic ground and from China. We hope this bilingual webzine will be a real 'bridge' between the English speaking world and China.

4 Conclusions

After two years hard work, the impact of uiGarden webzine on the information exchange is dramatic. Numerous western practitioners including usability gurus like

Don Norman, Jared Spool, Aaron Marcus, etc. have featured on the site, given Chinese practitioners and researchers the most fresh information from the western world. At the same time, the webzine has helped more and more Chinese practitioner introducing their practice in China in English for peers in the west to understand what's happening in China. At the time when this paper is written, there are more than 1000 subscribers to the webzine from all over the world.

There is still much to learn from the project. We believe that as more effective communication between the east and the west, western designers will gain more ideas when designing for global products. On the other hand, learning usability theories and methods without language barriers will also bring more people in China to understand and accept the concept and enable the Chinese usability and design industries to keep up with the world.

References

1. http://www.visualize.uk.com/bcshci/conslist/searchresults.asp
2. http://www.hcibib.org/hci-sites/COMPANIES.html
3. Dong, J., Fulimin, S.G.: Human-Computer Interaction:User Centered Design and Evaluation. Tstinghua University publishing, p. 1–3 (2003)
4. Smith, A., Yetim, F.: Global human-computer systems: cultural determinants of usability. Special Issue Editorial Interacting with Computers 16(1), 1–5, Elsevier (2004)
5. Smith, A., Dunckley, L.: Using the LUCID method to optimize the acceptability of shared interfaces. Interacting with Computers, vol. 9(3), pp. 335–345
6. Smith, A., Joshi, A., Liu, Z., Bannon, L., Gulliksen, J., Li, C.: Institutionalizing HCI in Asia. Interact 2007 (2007)
7. Li, C., Lisney, E., Liu, S.: Two Languages, Two Forums and a Cultural Exchange, EVA, London (2005)
8. http://www.chinaui.com/text/t1.asp?id=362
9. Li, C.: The Practice of a Bilingual Webzine. UPA 2005 (2005)

Designing "Culture" into Modern Product: A Case Study of Cultural Product Design

Rungtai Lin[1], Ming-Xian Sun[1], Ya-Ping Chang[1], Yu-Ching Chan[1], Yi-Chen Hsieh[2], and Yuan-Ching Huang[2]

[1] Department of Crafts and Design, Professor, National Taiwan University of Arts
59, Section 1, Ta-Kung Road, Pan-Chiao City, 220, Taipei, Taiwan
rtlin@mail.ntua.edu.tw
[2] Graduate School of Design, Chang Gung University, Tauyuan, Taiwan
rolabow2@gmail.com

Abstract. "Culture" plays an important role in the design field, and "cross cultural design" will be a key design evaluation point in the future. Designing "culture" into modern product will be a design trend in the global market. Obviously, we need a better understand of cross-cultural communication not only for the global market, but also for local design. While cross-cultural factors become important issues for product design in the global economy, the intersection of design and culture becomes a key issue making both local design and the global market worthy of further in-depth study. The importance of studying culture is shown repeatedly in several studies in all areas of technology design. Therefore, this study focuses on the analysis of cultural meaning, operational interface, and the scenario in which the cultural object is used. This paper establishes a cultural product design model to provide designers with a valuable reference for designing a successful cross-cultural product.

Keywords: cross cultural design, cultural difference, Taiwan aboriginal culture.

1 Introduction

We now live in a small world with a global market. While the market heads toward "globalization", design tends toward "localization." So we must "think globally" for the market, but "act locally" for design. Designing local features into a product appears to be more and more important in the global market where products are losing their identity because of the similarity in their function and form [3]. Cultural features then are considered to be a unique character to embed into a product both for the enhancement of product identity in the global market and for the fulfillment of the individual consumer's experiences [10]. The increasing emphasis on localized cultural development in Taiwan demonstrates an ambition to promote the Taiwanese style in the global economic market as for example, the use of aboriginal music from the Amis tribe at the 1996 Olympic Games which brought that form of music to the global arena. Additionally, martial art movies from Bruce Lee to Jacky Chan to the Oscar-winning movie director Ang Lee, have promoted recognition of the Taiwanese culture at the international level [5,2].

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 146–153, 2007.
© Springer-Verlag Berlin Heidelberg 2007

The "Economic Miracle" of Taiwan was made possible by the hard work of Taiwanese "industriousness and thrift". In the OEM (Original Equipment Manufacture) era, Taiwan's manufacturers reduced costs to produce "cheap and fine" products to be successful in the global market. After 1980, Taiwan enterprises began to develop ODM (Original Design Manufacture) patterns to extend their advantages in OEM manufacturing. Recently, product design in Taiwan has stepped into the OBM (Original Brand Manufacture) era. In addition, cultural and creative industry have already been incorporated into "Challenge 2008: National Development Grand Plan", demonstrating the government's eagerness to transform Taiwan's economic development by "Branding Taiwan" using "Taiwan Design" based on Taiwanese culture and aboriginal culture [10].

In the global market - local design era, connections between culture and design have become increasingly close. For design, cultural value-adding creates the core of product value. It's the same for culture; design is the motivation for pushing cultural development forward [4]. Therefore, based on the "Taiwan experience", this article intended to propose a cultural product design model and examples illustrating how to transfer "Taiwan culture" to design elements, and design these cultural features into modern products to reinforce their design value. Results presented herein create an interface to look at the way Taiwan designers communicate across cultures, as well as the interwoven experience of design and culture in the design process.

2 Taiwan Aboriginal Culture

Taiwan is a multi-cultural blend of traditional Chinese with significant East Asian influences including Japanese and such Western influences as American, Spanish and Dutch. Over time, Taiwan gradually developed its own distinct culture, mostly from a variation of Chinese culture coming from southern China. Of course, the Taiwanese aboriginals also have a distinct culture. The Taiwan Aborigines are the original inhabitants of Taiwan, residing there long before the Chinese immigrants came from Mainland China to Taiwan during the Ching Dynasty from mid-17th century. They originally derive from the Austronesian Language Family. According to anthropological studies of physical features, customs, thoughts, language, and verbal narrative history, they are identified with twelve tribes: Atayal, Saisiat, Bunun, Tsou, Thao, Paiwan, Rukai, Amis, Puyuma, Tao, Kavalan, and Truku [1,7,8,9].

Among these 12 tribes, there are different customs and material cultures from one tribe to the other, particularly in different geographical environments. For instance, the culture of tribes located near the ocean reflects their fishing based living, while mountain tribe culture develops from a dependence on hunting [18,19]. However, in general, traditional tribes have self-sufficient societies which are dependent on agriculture, fishing, hunting and animal husbandry. It is very important to study Taiwan Aboriginal totem art in order to understand their culture because of their lack of a written language [9,13,14,16]. The totems appeared on textiles and sculptures and can illustrate the culture itself. With their beautiful and ancestral visual arts and crafts, Taiwan aboriginal cultures should have great potential for enhancing product design value thus increasing recognition in the global market. By enhancing the original meaning and images of Taiwan Aboriginal cultures and taking advantage of new

production technology, designers in Taiwan are trying to transform aboriginal cultural features into modern products and fulfill the needs of the contemporary consumer market.

3 A Cultural Product Design Model

Cultural product design is a process of rethinking or reviewing cultural features and then to redefining the process in order to design a new product to fit into society and satisfy consumers with via culture and esthetic. Using Cultural features to add extra value to a product can not only benefit economic growth but also promote unique local culture in the global market [6,15]. Therefore, how to transfer cultural features into a cultural product becomes a critical issue. A cultural product design model which facilitates understanding of cultural product design is proposed as shown in Figure 1 [11].

In Figure 1, the cultural product design model consists of three main parts: conceptual model, research method, and design process. The conceptual model focuses on how to extract cultural features from cultural objects and then transfer these features to a design model to design cultural products. The research method consists of three steps: identification, translation and implementation, to extract cultural features from original cultural objects (identification), transfer them to design information and design elements (translation), and finally design a cultural product (implementation).

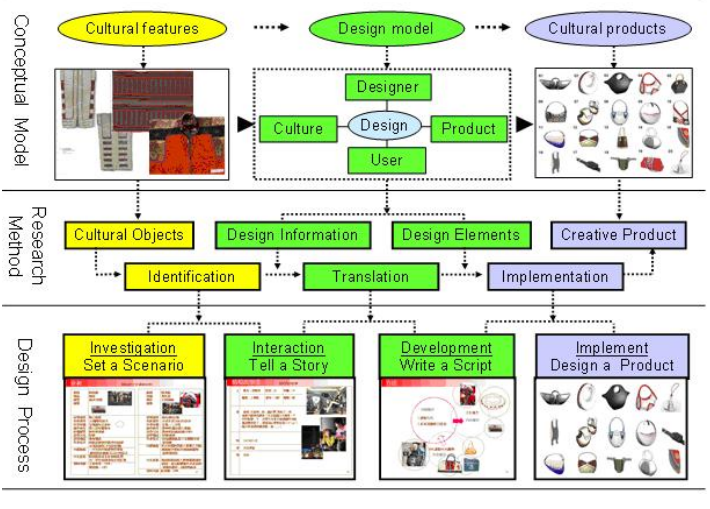


Fig. 1. Cultural product design model

Based on the cultural product design model, the cultural product was designed using scenario and story-telling approaches. In a practical design process, four steps are used to design a cultural product, namely, investigation (set a scenario),

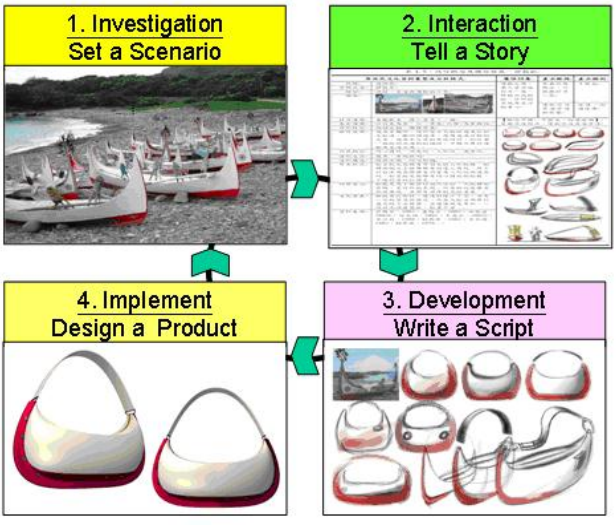


Fig. 2. The cultural product design process

interaction (tell a story), development (write a script), and implementation (design a product) as shown in Figure 2.

(1) Investigation/set a scenario: the first step is to find the key cultural features from the original cultural object and to set a scenario to fit the three levels: outer ‘tangible’ level, mid ‘behavioral’ level, inner ‘intangible’ level. Based on the cultural features, the scenario should consider the overall environment such as economic issues, social culture, and technology application. This step tends to analyze the cultural features in order to determine the key cultural features to for representing the product.

(2) Interaction/tell a story: based on the previous scenario, this step focuses on a user-based observation to explore the social cultural environment in order to define a product with cultural meaning and style derived from the original cultural object. Therefore, some interactions should be explored in this step including interaction between culture and technology, dialogue between users and designers, and understanding the user’s need and cultural environment. According to the interaction, a user-centered approach was used to describe the user need and the features of the product by a story-telling.

(3) Development/write a script: this step is the concept development and design realization. The purpose of this step is to develop idea sketch in text and pictograph form through the access of scenario and story. During this process, modification of the scenario and story might occur in order to transform the cultural meaning into a logically correct cultural product. This process provides a way to confirm or clarify the reason why a consumer needs the product and how to design the product to fulfill the users’ need.

(4) Implement/design a product: this step deals with identified cultural features and the context of cultural products. At this stage, all cultural features should be listed in a matrix table which will help designers check the cultural features in the design process. In addition, the designer needs to evaluate the product features, product meaning, and the appropriateness of the product. The designer may make changes to the prototype based on the results from the evaluation, and implement the prototype and conduct further evaluations.

4 Design Case Study of Scenario and Story-Telling Approach

Taking Tao culture as an example, the Tao people are a Taiwan aboriginal people who are native to the tiny outlying Orchid Island. The Tao people are traditionally good at making canoes. The Pin-Ban boat shown in Figure 2 is a symbol of their tribe. The Tao people live by fishing and usually bring the holy dagger with them while fishing. Figure 3 shows the final cultural product designed from the Tao's Pinban boat and holy dagger. The scenario is that Tao people ride in their Pin-Ban boat and bring their holy dagger to protect them and sail to the ocean for fishing. Based on the scenario, the Pin-Ban boat was transformed into a modern bag and the holy dagger into a knife-like modern alarm. In modern society, one can imagine a pretty woman holding the modern bag and bringing the modern alarm to protect her while walking down the street as matching the previous scenario of Tao people fishing with their Pin-Ban boat and holy dagger [11].



Fig. 3. The cultural product from Pin-Ban boat and holy dagger

Based on the cultural product design model, figure 4 showed how to transfer the original object -- 'Pottery-pot' from Paiwan tribe to design a modern bag. Different cultures use textile containers designed for their own storage and transportation needs.

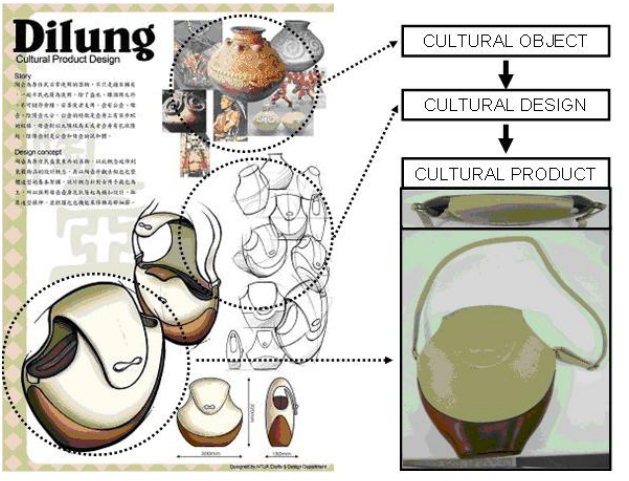


Fig. 4. The modern bag deigning from the pottery pot



Fig. 5. The process of designing modern bags from cultural objects

Unlike bags or containers made from rigid materials such as clay or glass, textile containers offer flexibility of use by adapting to whatever item they are carrying. Figure 5 shows how to use the Taiwan aboriginal garment as the original cultural objects to design modern bags. In addition, figure 6 demonstrated the cultural features extracted from Taiwan aboriginal garment culture and transformed into modern bag design.

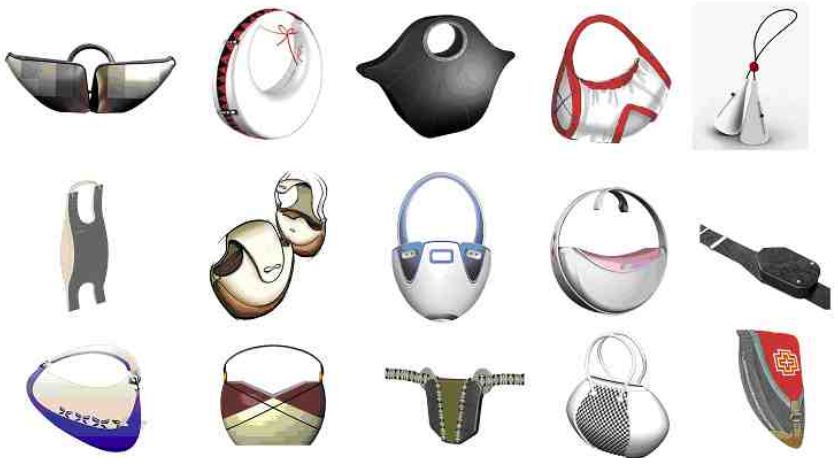


Fig. 6. Various bags from the Taiwan aboriginal garment culture

5 Conclusion

In recent years, if you have seen BENQ, ASUS, DUCK, or other design teams on the prize list of international design contests and wondered if there were Taiwanese designer behind them, you would have been correct. Recently, Design teams from Taiwan have become regular winners of international design contests. The capability of Taiwan's industrial design is coming to the fore in of the global design arena through international design contests. Therefore, based on "Taiwan culture", the purpose of this article is to show how "Branding Taiwan" into "Taiwan Design" can reinforce design value upwards. For "Branding Taiwan", we must think about how to use the Taiwan culture. For the "Taiwan Design", we need to design "Taiwan cultural features" into modern products.

The aboriginal garment culture provides a good example of applying cultural features to design while still retaining a meaningful cultural value. This paper demonstrated the cultural features of aboriginal garment culture and how to transform those features into a new cultural product design which can fit into the contemporary market. Cultural products, hence, can extend the heritage and traditional values of Taiwan Aboriginal Culture to the consumer and increase the sense of spiritual essence in human life. Perhaps the best way to extend Taiwan Aboriginal Culture is to promote it in consumers' daily lives through impressions made by the use of products such as garments, crafts, decorations, utensils, furniture, ornaments, and packages whose designs are based on that culture.

Acknowledgments. I gratefully acknowledge the support for this research provided by the National Science Council under Grant No. NSC-94-2422-H-144-001, NSC-94-2422-H-144-003 and NSC 95-2422-H-144 -003. The author wishes to thank the various students and colleagues who have contributed to this study over the years, specially, Dr. J. G. Kreifeldt and Mr. T.U. Wu.

References

1. Chang, H.G.: (2006) Taiwan aboriginal research website, http://www.lib.nthu.edu.tw/library/hslib/subject/an/native.htm
2. Cheng, H.: The application of Taiwanese aboriginal culture in product design. Chang Gung University, Industrial Design Department, Master Thesis (In Chinese) (2005)
3. Handa, R.: Against arbitrariness: architectural signification in the age of globalization. Design Studies 20, 363–380 (1999)
4. Ho, M.C., Lin, C.H., Liu, Y.C.: Some speculations on developing cultural commodities. Journal of Design 1, 1–15 (in Chinese) (1996)
5. Hsu, C.-H.: An application and case studies of Taiwanese aboriginal material civilization confer to cultural product design, Chang Gung University Industrial Design Department, Master Thesis (In Chinese) (2004)
6. Lee, K.P.: Design methods for cross-cultural collaborative design project. In: Redmond, J., Durling, D., De Bono, A. (eds.) Proceedings of Design Research Society International Conference. Paper #135, DRS Futureground, Monash University, Australia (2004)
7. Lee, S.L.: Garments culture of Taiwan Aborigines. Historical Objects 87, 14–28 (In Chinese) (2000)
8. Lee, R.-K.: The immigration of Taiwan southern island tribes, Charng-Ming Culture Publisher, Taipei (In Chinese) (1997)
9. Lin, J.-C.: Taiwan aboriginal art?Field study, Art book, Taipei (In Chinese) (2002)
10. Lin, R.: Creative learning model for cross cultural product. Art. Appreciation 1(12), 52–59 (2005)
11. Lin, R.: Scenario and story-telling approach in cross cultural design. Art. Appreciation 2(5), 4–10 (2006)
12. Lin, S.G.: The study of pottery pattern and decorative skills in application. Taiwan Crafts 10, 23–48 (In Chinese) (2002)
13. Liou, C.W.: Taiwan aboriginal culture art, Lion publisher, Taipei (In Chinese) (1979)
14. Liu, M.C.: Culture and art of the Formosan aboriginal, Hsiung-Shih Art Book, Taipei (In Chinese) (1982)
15. Leong, D., Clark, H.: Culture -based knowledge towards new design thinking and practice - A dialogue. Design Issues 19(3), 48–58 (2003)
16. Shu, G.M.: The culture and art in Rukai tribe, Rice Publishing, Taipei (in Chinese) (1998)
17. Taiwan aborigines art studio (2006) http://www.sandiman-sct.idv.tw/new_page_1.htm
18. Wu, T.Y., Cheng, H., Lin, R.: The study of cultural interface in Taiwan Aboriginal Twin-Cup. HCI INTERNATIONAL 2005, 22-27 July, 2005, Las Vegas - Nevada USA. Paper-ID1712,CD-ROM Format (2005)
19. Wu, T.Y., Hsu, C.H., Lin, R.: The study of Taiwan aboriginal culture on product design. In: Redmond, J., Durling, D., De Bono, A. (eds.) Proceedings of Design Research Society International Conference. Paper #238, DRS Futureground, Monash University, Australia (2004)

Digital Archive Database for Cultural Product Design

Rungtai Lin[1], Ricer Cheng[2], and Ming-Xian Sun[3]

[1] Department of Crafts and Design, National Taiwan University of Arts
59, Section 1, Ta-Kung Road, Pan-Chiao City, 220, Taipei, Taiwan
[1] rtlin@mail.ntua.edu.tw, [2] vigital@ms57.hinet.net,
[3] puma_fox@yahoo.com.tw

Abstract. The purpose of this paper is to build a digital archive database for Taiwanese people to learn Taiwanese culture through the internet and e-learning environment. This study will be completed in three steps. Firstly, the paper is to explore the meaning of cultural objects and to extract the cultural features from Taiwanese culture; especially, Taiwan ordinary garment cultures. Then, a protocol of information-exchanging is used to analyze the cultural features, and to combine the images with the text using a standardization of digital images. Finally, a digital archive database is established and a friendly interface is designed for users. Results are presented here providing the users with a digital archive database to learn Taiwan local cultural features.

Keywords: Digital archive, database, cultural design, Taiwan aboriginal culture.

1 Introduction

Taiwan is a multi-culture blend of traditional Chinese with significant East Asian influences including Japanese and such Western influences as American, Spanish and Dutch. Over time, Taiwan gradually developed its own distinct culture, mostly from a variation of Hoklo culture from Southern China. Of course, the Taiwanese aboriginals also have distinct cultures [1], [23]. Designing local features into products appears to be more and more important in the global market, where products are losing their identity because of their similarity in function and form. Cultural features then are considered to be unique characters to embed into a product both for the enhancement of its identity in the global market and for the fulfillment of the individual consumer's experiences [4], [26], [27]. The increasing emphasis on localized cultural development in Taiwan demonstrates an ambition to promote the Taiwanese style in the global economic market [3], [6].

Using local features in design fields as a strategy to create product identity in the global market, the designer has noted the importance of associating products with cultural features in order to enhance the product value. At this point, the field of Industrial Design has played an important role in embedding the cultural elements into products and in increasing the cultural value in the global competitive product market. Therefore, designing a product with local features in order to emphasis its cultural value has become a critical issue in the design process [15], [24], [25].

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 154–163, 2007.
© Springer-Verlag Berlin Heidelberg 2007

"Culture" plays an important role in the design field, and "cross-cultural design" will be a key design evaluation point in the future. Designing "culture" into products will be a design trend in the global market. Obviously, we need a better understanding of cross-cultural communications not only for the global market, but also for local design. While cross-cultural issues become important for product design in the global economy, the intersection of design and culture become a key issue making both local design and the global market worthy of further in-depth study. The importance of studying culture is shown repeatedly in several studies in all areas of technology design [5], [15].

The *Linnak*, literally "twin-cup" in the Paiwan language, was chosen as the cultural object for this paper and its appearance, usability, and cultural meaning were studied. The *Linnak* is a special container for drinking wine used in the traditional tribal wedding ceremony and requires two people to manipulate the object smoothly for the drinking task. There are social meanings, ergonomic concerns and the functional achievement associated with this cultural object. Therefore, this study focuses on the analysis of cultural meaning, operational interface, and the scenario in which this object is used. Then, an archive database model was established to provide designers with a valuable reference for designing a successful cultural product. Results presented herein create an interface for examining the way designers communicate across cultures as well as the interwoven experience of design and culture in the design process [15].

2 Culture and Cultural Design Features

Culture has been called "the way of life for an entire society." [5], [20]. It generally refers to patterns of human activity and the symbolic structures that give such activity significance. Different definitions of "culture" reflect different theoretical bases for understanding, or criteria for evaluating, human activity. Based on linguistic, anthropological, and sociological studies, culture has been described as that which deals with the result of the evolutionary process in human civilization that involves language, customs, religion, arts, thought and behavior. From the design point of view, Lee [8] proposed a culture structure with multi-layers including 'artifact', 'value', and 'basic assumptions' which identified key design attributes including 'functional', 'aesthetic', and 'symbolic'. Leong and Clark [20] developed a framework for studying cultural objects distinguished by three special levels: the outer 'tangible' level, the mid 'behavioral' level, and the inner 'intangible' level.

Based on previous studies [8], [15], [16], [20], [25], a framework for studying cultural objects is summarized as shown in Figure 1 (Lin, 2005). As shown in Figure 1, culture can be classified into three layers: (1) physical or material culture -- including food, garments, and transportation related objects, (2) social or behavioral culture-including human relationships and social organization, and (3) spiritual or ideal culture -- including art and religion. These three culture layers can be fitted into Leong's three culture levels given above. While cultural objects can be incorporated into cultural design, three design features are identified as follows: (1) the inner level conta-inning special content such as stories, emotion, and cultural features, (2) the mid

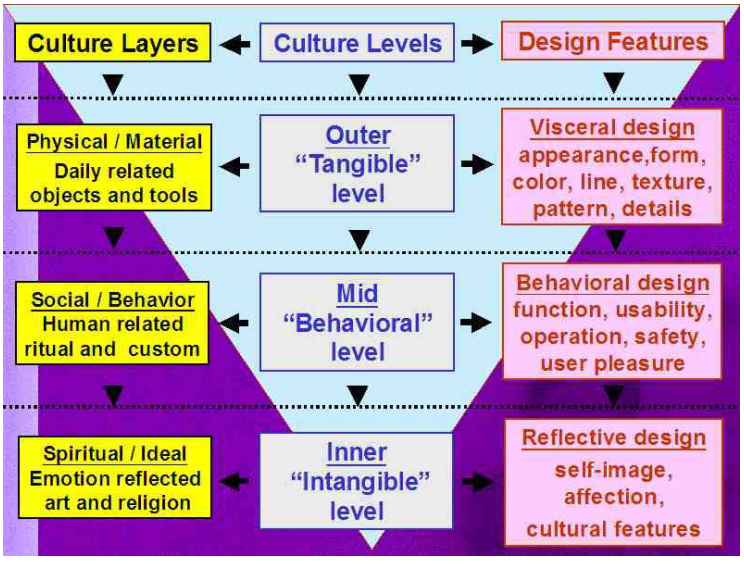


Fig. 1. Three layers and levels of cultural objects and design features

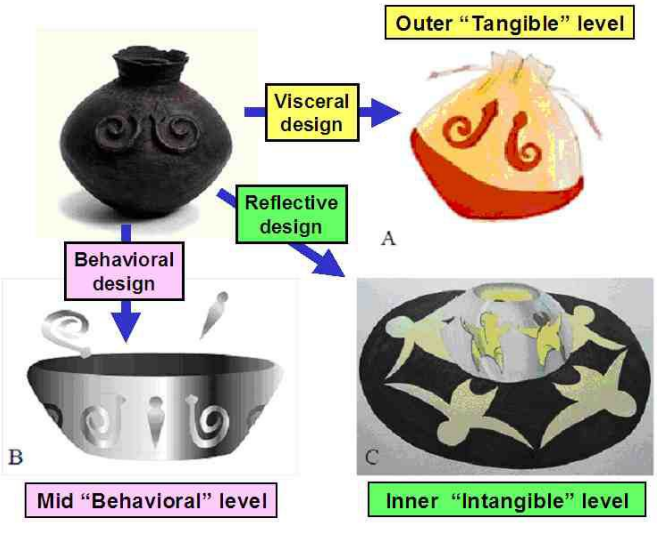


Fig. 2. Three levels of a cultural product and its design features

level containing function, operational concerns, usability, and safety, and (3) the outer level dealing with colors, texture, form, decoration, surface pattern, lines quality, and details [20].

Taking Taiwan aboriginal culture as an example, Figure 2 illustrates the application of the three levels of a "Pottery-pot" from Paiwan Tribe to design a cultural product [15]. The three levels of the cultural object can be mapped into three

levels of design features: visceral design, behavioral design and reflective design, in Norman's book -- 'Emotional Design.' Visceral design concerns the appearance of a cultural object, and transforms its form, textures, and pattern into a new product. The visceral design feature is where appearance matters and first impressions are formed. The behavioral design level is about use, function, performance and usability of a cultural object. The behavioral design feature is the key to a product's usefulness. Reflective design concerns feeling, emotion, and cognition of a cultural object. The reflective design feature is the most vulnerable to variability through culture, experience, education, and individual difference.

3 Cultural Levels and Design Features

Among Taiwan Aboriginal tribes, each tribe has a unique culture and style, which can be identified simply from its sculptures, textile fabrics, webbing, leather craft, and pottery. Furthermore, the tribe can be identified through the applied functions of architectures, daily life objects, tools, ceremonial equipment, weapons, and decorations. Upon investigation of the equipments and tools of the 12 Taiwan Aboriginal tribes, the Linnak from Paiwan tribe displays a remarkable usefulness as a cultural resource [2], [7], [9], [10], [11], [12]. The unique shape of the Linnak shows its cultural meaning, usability, and beauty. Therefore, the Linnak is used as an example to demonstrate its cultural features in product design application.



Fig. 3. The Linak–twin cup from Paiwan culture

The Linnak, a twin cup, is a very typical object in Paiwan culture. In the Paiwan language, "Linnak" represents the value of connection to Paiwan traditional culture. The Linnak is carved from one piece of wood and usually has two cups with one handle on each side as shown in Figure 3. Taiwan Aboriginal people often drank rice

wine and considered it to be a holy event. Linnak represents their drinking culture and the meaning of drinking [13], [17], [28], [19], [22]. In ancient times, therefore, they developed a variety of drinking containers and each container has an associated special events and meaning. For example, one cup can be used only in special events by the chief of Paiwan tribe; two-cups or three-cups was normally used in wedding or festival ceremonies to enhancement the close relationship of drinkers and to increase warm feelings as shown in Figure 3.

There are social meanings, ergonomic concerns and the functional achievement associated with this cultural object. To provide an ideal drinking cup at a wedding, both the social and operational interfaces of the "*Linnak*" need to be well-designed. The design features of Linnak has been identified with three levels of cultural features: (1) outer-level focus on the Linnak formation which is associated with material, colors, texture, and pattern; (2) Mid-level focus on the consumer behavior and the scenario in which people will use the Linnak in what kinds of occasion; and (3) inner-level focus on the symbolic meaning of the Linnak [8], [15], [20], [21].

4 Cultural Feature Transformation Model

Based on the previous cultural levels and design feature, a cultural features transformation model was proposed which consists of three main parts as shown in Figure 4, including cultural features, design model, and cultural product. The conceptual model focuses on how to extract cultural features from cultural objects and then transfer these features to the design model to design cultural products. The research method consists of three steps; identification, translation and implementation, to extract cultural features from original cultural objects (identification), transfer them to design information and design elements (translation), and finally design a cultural product (implementation). The research method is described as follows:

(1) Identification phase: The cultural features were identified from original cultural objects including the outer level of colors, texture, and pattern, the mid level of function, usability, and safety, and the inner level of emotion, cultural meaning, and stories. The designer used the scientific method and other methods of inquiry and hence is able to obtain, evaluate, and utilize design information from the cultural objects.

(2) Translation phase: A translation phase has translated the design information to design knowledge within a chosen cultural object and has achieved some depth and experience of practice in these design features. At the same time, the designer is able to relate this design knowledge to design problems in modern society, and has an appreciation for the interaction between culture, technology, and society.

(3) Implementation phase: An implementation phase possesses the design knowledge associated with the cultural features, the meaning of culture, an aesthetic sensibility, and the flexibility to adapt to various designs. At this time, the designer has knowledge of cultural objects and an understanding of the spectrum of culture and value related to the cultural object. The designer combines this knowledge with his strong sense of design to deal with design issues, and to employ all of the cultural features in designing a cultural product.

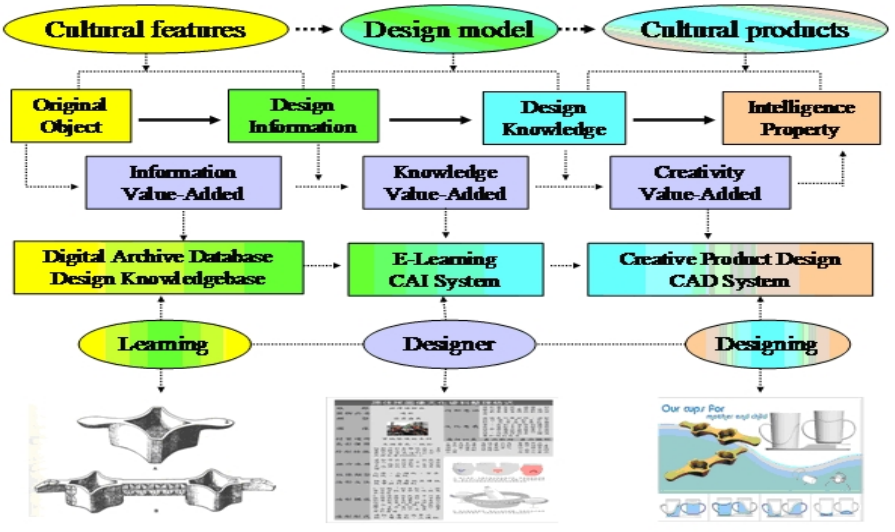


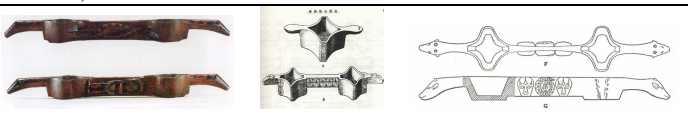
Fig. 4. The cultural product design model consists of three main parts

5 Applications of Digital Archive Database

The application of cultural features is a powerful and meaningful approach to product design. Consumers nowadays require a design which is not only functional and ergonomic, but which also stimulates emotional pleasure. The results of studying the Linnak shown in Table 1 demonstrate that cultural features are a valuable element to embed into a product to emphasize its value or meaning.

Table 1 shows an example of cultural features and design attributes of the Linnak from the Paiwan tribe. Based on the cultural feature transformation mentioned above, table 1 illustrates three levels of Linnak information in Paiwan culture: physical appearance, interactive behaviors, and spiritual meaning. As mentioned above, material culture has illustrated a strong impact in design application in terms of visual image stimulation and symbolization of the objects. This study first focused on data collection from the Paiwan in the aspects of physical, material, customs, ceremonies, and spirituality among the objects. The collected data as shown in Table 1 was then matched to the different items based on tribe, name of object, type, image, material, color, appearance, usability, pattern, form grammar, form structure, form style, inner content, and original resource. These items covered three levels of cultural characteristics and basic information such as imagery icon, tribe, and name. We propose that this information will serve as a reference for designers during the product design phase (Cheng, 2005; Hsu, 2004; Chen, 1961; Liu, 1982).

Table 1. Cultural features of Linnak Container from the Paiwan tribe

Object	Linnak or twin cup
Type	Drinking container
Tribe	Paiwan, Rukai
Picture	
Material	Wood
Color	Natural wood color or painting with colors
Pattern	Figure, human-head, long-hooded pit viper pattern, Deer pattern
Principle of formation	1. Embossment on handles with a variety of patterns. 2. Total length from 43cm to 91cm, pitch from 29cm to 42cm, and cup capacity from 300c.c. to 600c.c. 3. Single cup with a rectangular column shape and handle on both sides. 4.The Linnak contains two rectangular column cups, a beam bridge in between and a handle on both sides.
Classification	Twin-cup, Single-cup and Tri-cup.
Operation	Two drinkers are required to hold the handles with left and right hand on each when drinking alcohol.
Using Scenario	1. Single cup is created only for the chief to drink liquid in the Paiwan tribe. Sometimes it was used to contain rice alcohol and to reward a hunting hero for demonstrating valor. 2. The twin -cup was created for use in wedding ceremonies where the bride and groom were required to drink alcohol together. 3. The Tri-cup was created for the groom and bride and chief (a witness), who stands between groom and bride to drink alcohol together which represents the greatest honor and wish for the couple.
Cultural content	1. The long- hooded pit viper or ancient figure pattern on the cup enhances the value of cup. 2. The twin-cup was mostly used in ritual or festival ceremonies to demonstrate a warm and harmonious spirit. 3. To drink with the twin cup represented the commitment in love between a male and female in tradition culture. 4. Drinking together represents eternal friendship.

According to Table 1, a digital archive database was built to help to understand both the hard and soft contents of the cultural object. A process of building a digital archive database included six steps (figure 5): select the cultural object, deal with the image, collect the information, transfer to the design knowledge, format the related information, and build the database. In addition, a friendly interface was provided to designer for accessing the database easily as shown in Figure 6, 7, 8. Some cultural products design are shown in Figure 9 and 10 on the basis of the digital archive database.

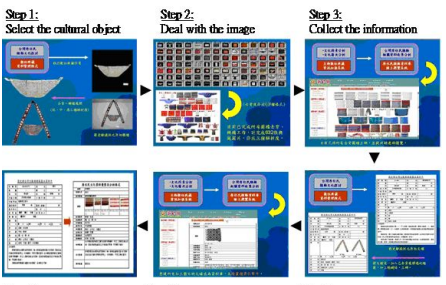


Fig. 5. Process of building the database



Fig. 6. Interface of the database

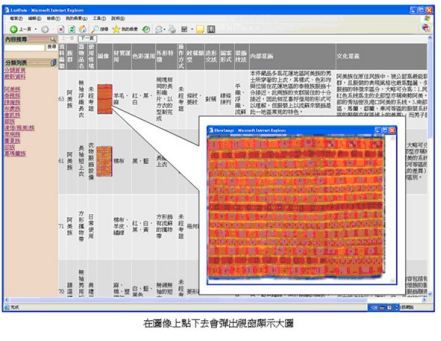


Fig. 7. Interface for referring the pattern

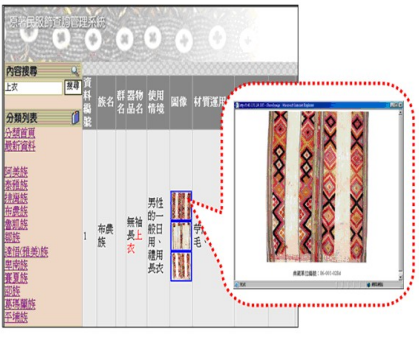


Fig. 8. The detail of the pattern

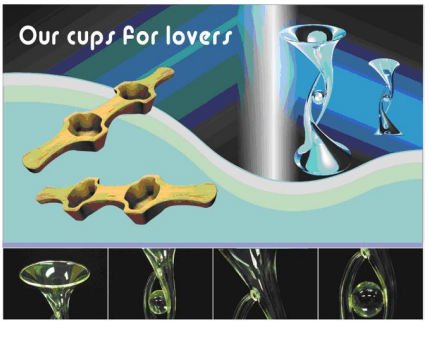


Fig. 9. Our cups for lovers

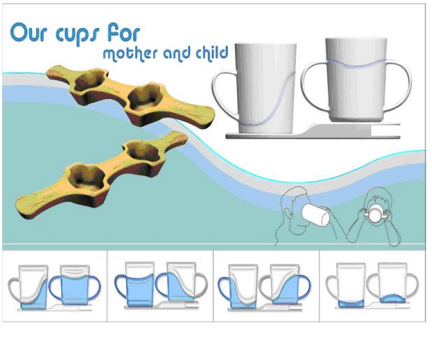


Fig. 10. Our cups for mother and child

6 Conclusion and Suggestion

It is noted that the beauty of Taiwan Aboriginal culture and art demonstrates a great potential for enhancing the design value of modern consumer products. With beautiful and primitive visual art and crafts, Taiwan aboriginal culture should have great

potential to enhance the design value, and to be recognized in the global market. Evidence shows that the perspective of Taiwan local culture in design will undoubtedly become crucial cultural elements in future design applications. Therefore, a cultural feature transformation model was proposed for transforming Taiwan aboriginal culture features into modern product design.

In this study, the Linnak demonstrates the value of Aboriginal culture in design. Based on the cultural feature transformation model, the Linnak has been identified with three levels of cultural interfaces and used as an example to demonstrate how to build a digital archive database. The Linnak is a typical cultural object which can be transformed into a contemporary design for the current consumer market. The idea of sharing in the design of the Linnak is valuable for enhancing usage in our daily life. However, the contemporary consumer market may need a new form of the Linnak suitable for a modern environment. In other words, a transformation of the Linnak is necessary for the modern market.

For future study, we suggest field investigations and interviews with Taiwanese Aboriginal people in addition to literature review in order to accurately understand Aboriginal culture and art so as to avoid incorrect interpretation of the culture when transforming cultural features into modern product design. In addition, a detailed design process needs to be developed in order to provide designers with the specified procedures for designing cultural products in the future.

Acknowledgments. I gratefully acknowledge the support for this research provided by the National Science Council under Grant No. NSC-94-2422-H-144-001, NSC-94-2422-H-144-003 and NSC 95-2422-H-144 -003. The author wishes to thank the various students and colleagues who have contributed to this study over the years, specially, Dr. J.G. Kreifeldt and Mr. T.U. Wu.

References

1. Chang, H.G.: (2006) Taiwan aboriginal research website http://www.lib.nthu.edu.tw/library/hslib/subject/an/native.htm
2. Chen, C.-L.: Woodcarving of the Paiwan group of Taiwan, SMC Publishing INC., Taipei (In Chinese) (1961)
3. Cheng, H.: The application of Taiwanese aboriginal culture in product design. Chang Gung University, Industrial Design Department, Master Thesis (In Chinese) (2005)
4. Handa, R.: Against arbitrariness: architectural signification in the age of globalization. Design Studies 20, 363–380 (1999)
5. Ho, M.C., Lin, C.H., Liu, Y.C.: Some speculations on developing cultural commodities. Journal of Design 1, 1–15 (In Chinese) (1996)
6. Hsu, C.-H.: An application and case studies of Taiwanese aboriginal material civilization confer to cultural product design, Chang Gung University Industrial Design Department, Master Thesis. (in Chinese) (2004)
7. Hu, G.Y.: The culture of Saisiat: tradition and evolution, Ministry of the Interior research report (In Chinese) (1996)
8. Lee, K.P.: Design methods for cross-cultural collaborative design project. In: Redmond, J., Durling, D., De Bono, A. (eds.) Proceedings of Design Research Society International Conference. Paper #135, DRS Futureground, Monash University, Australia (2004)

9. Lee, S.L.: Garments culture of Taiwan Aborigines. Historical Objects 87, 14–28 (in Chinese) (2000)
10. Lee, T.-C.: The study of Taiwan Aboriginal Paiwan Totan in application, National Taiwan Normal University, Master Thesis. (In Chinese) (2000)
11. Lee, Y.-Y.: Taiwan aboriginal society and culture, Linking Book, Taipei (In Chinese) (1982)
12. Lin, J.-C.: Taiwan aboriginal art?Field study, Art book, Taipei (In Chinese) (2002)
13. Lin, M.H., Huang, C.C.: The logic of the figurative expressions and cognition in design practices. Journal of Design 7(2), 1–21(In Chinese) (2002)
14. Lee, R.-K.: The immigration of Taiwan southern island tribes, Charng-Ming Culture Publisher, Taipei (In Chinese) (1997)
15. Lin, R.T.: Creative learning model for cross cultural product. Art. Appreciation 1(12), 52–59 (2005)
16. Lin, R.T.: Scenario and story-telling approach in cross cultural design. Art. Appreciation 2(5), 4–10 (2006)
17. Lin, S.G.: The study of pottery pattern and decorative skills in application. Taiwan Crafts 10, 23–48 (In Chinese) (2002)
18. Liou, C.W.: Taiwan aboriginal culture art, Lion publisher, Taipei (In Chinese) (1979)
19. Liu, M.C.: Culture and art of the Formosan aboriginal, Hsiung-Shih Art Book, Taipei. (In Chinese) (1982)
20. Leong, D., Clark, H.: Culture -based knowledge towards new design thinking and practice - A dialogue. Design Issues 19(3), 48–58 (2003)
21. Moalosi, R., Popovic, V., Hudson, A.: Socio-cultural factors that impact upon human-centered design in Botswana. In: Redmond, J., Durling, D., De Bono, A. (eds.) Proceedings of Design Research Society International Conference. Paper #716, DRS Futureground, Monash University, Australia (2004)
22. Shu, G.M.: The culture and art in Rukai tribe, Rice Publishing, Taipei (in Chinese) (1998)
23. Taiwan aborigines art studio (2006) http://www.sandiman-sct.idv.tw/new_page_1.htm
24. Wu, T.Y., Cheng, H., Lin, R.: The study of cultural interface in Taiwan Aboriginal Twin-Cup. HCI INTERNATIONAL 2005, 22-27 July, 2005, Las Vegas - Nevada USA. Paper-ID1712,CD-ROM Format (2005)
25. Wu, T.Y., Hsu, C.H., Lin, R.: The study of Taiwan aboriginal culture on product design. In: Redmond, J., Durling, D., De Bono, A. (eds.) Proceedings of Design Research Society International Conference. Paper #238, DRS Futureground, Monash University, Australia (2004)
26. Yair, K., Press, M.: Crafting competitive advantage: crafts knowledge as a strategic resource. Design Studies 22, 377–394 (2001)
27. Yair, K., Tomes, A., Press, M.: Design through marking- crafts knowledge as facilitator to collaborative new product development. Design Studies 20(6), 495–515 (1999)

Cross-Cultural Understanding of Content and Interface in the Context of E-Learning Systems

Abdalghani Mushtaha and Olga De Troyer

Vrije Universiteit Brussel, Department of Computer Science
Research Group WISE, Pleinlaan 2, 1050 Brussels, Belgium
{abmushta,Olga.DeTroyer}@vub.ac.be

Abstract. This paper describes a comparative study in understanding content and interface in the context of e-learning systems by using anthropologists' and designers' cultural dimensions. The purpose was to determine the differences between Belgian and Palestinian audiences, and to find the most important cultural dimensions to use for localizing / internationalizing e-learning systems. Results indicate differences in culture between the two groups, but not as much as expected. The outcome shows similar preferences, whilst others differ.

Keywords: e-learning, Web design, Cross-cultural dimensions, localization, Internationalization.

1 Introduction

Cultural differences have been studied and researched by many anthropologists (e.g., Fons Trompenaars [17], Edward Hall [7], David Victor [18], Quincy Wright [14]) over recent years. Anthropologists have tried to divide and categorize the world, to find definitions for cultural values, and call them *cultural dimensions*.

Based on those cultural studies, much research done today tries to find a link between cultural dimensions and Website design. For example, Evers & Day, 1997 [5]; Marcus & Gould, 2000 [13]; Mahemoff & Johnston, 1998 [11]; Dormann & Chisalita, 2002 [4] have applied cultural dimensions to global interface design, tested users behavior, variations in understanding colors, icons, pictures, symbols, phrases etc. They found that culture affects users from different cultures, who may perceive the same Website in totally different ways. Some metaphors, navigation, text, graphical elements, or navigation might be misunderstood. Therefore, culture aspects should been taken into consideration during localization or Website development.

The well-known anthropologist Geert Hofstede [8] defines the culture as: "patterns of thinking, feeling, and potential acting that all people carry within themselves", and which he terms "mental programs". The source of these programs lies within the social environments in which people grew up and collected their life experiences. Culture affects who we are, how we think, how we behave, and how we respond to our environment. Above all, it determines how we learn. A person's cultural background is learned, not inherited and is made up of experiences gained when growing up in his or her culture. Therefore, cultural background is used in understanding the 'virtual' world on the screen. Moreover, Ruttenbur, spickler &

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 164–173, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Lurie [15] said "in E-learning, solutions facilitate the delivery of the right information and skills to the right people at the right time". Next to this, no one can deny that there are differences in what is acceptable from one culture or nation to another. Therefore, information needs to be presented in ways that allow the learner to interact with them, whilst taking into account cultural aspects.

In addition, according to Nielsen (1997) [9], many Web usability problems that arise are due to variations in behaviors and cultural differences. Such variations may be found in graphics elements, symbols, pictures, icons, text structure, idioms, help information, dates and numbers format, etc.

As students' expectations of education and universities, as well as the way students are educated and teachers communicate, may differ from one country to another, students with different cultural background may understand the same e-learning Website in different ways. This study highlights the differences between students from different cultural backgrounds in the way they understand text and graphics, use navigation and search, as well as in their attitudes towards e-learning Websites.

2 Purpose of study

A comparative study was carried out involving Palestinian and Belgian students to measure the understanding and acceptance of some pre-selected e-learning portals.

The purpose of this study was to explore and evaluate the influence of the user's cultural background on content and interface understanding in the context of an e-Learning system. Specifically, we were interested to determine the most important cultural dimensions, factors and issues that should be taken into consideration when designing an e-learning Website.

The theoretical frameworks that have been used to guide this study are the cultural dimensions of the anthropologists and systems designers. The 16 cultural dimensions considered in our experiment will be mentioned later on. Please note that most of these cultural dimensions have already been evaluated and tested by researchers in separate studies, to explore their influence on user-interface and systems design (e.g., Barber & Badre 1998 [3]; Ever & Day 1997 [5]; Marcus & Gould 2000 [13]; Stengers & De Troyer 2004 [16]). However, in these researches only some of the dimensions were considered while in our research we have used all the dimensions to find out which dimensions are the most important ones. In the study, we focused on the interface and the usability aspects as well as on the content.

3 Methodology - The Experiment

The subjects of the experiment were students from Palestine and Belgium. Questionnaires were sent out to 100 Belgian students and 100 Palestinian students. Responses were received from 42 Palestinian and 21 Belgian students, who were then requested to further participate in our experiment. The experiment was conducted by means of two different e-learning portals:

- IUG-WebCT[1]: a portal generated by the WebCT[2] tool. WebCT is a well-known tool for e-learning systems of higher education institutions. Thousands of colleges and universities in more than 70 countries worldwide use this tool. The target portal was developed for the Islamic University of Gaza. All students involved in the experiments were familiar with WebCT.
- CLC: Collaborative Learning Centre[3]: a portal that aims to provide a wide range of Web-based interactive courses. It gives opportunities to the teacher to fully control the courses taught and for students and teachers from different countries to meet and work with each other. None of the students involved in the experiment was used to work with CLC.

In our experiment we used a multi-method approach involving questionnaires, icon recognition exercises, hands-on observation, task scenarios and interviews. The questionnaire for this research study was 16 pages and divided into 3 main sections:

1. Participants' Characteristics: This section was used to collect
   - Demographic and personal details: age, gender, language background, etc.
   - Computer & Internet experience.
2. Cultural Evaluation: In this section, the 16 cultural dimensions were presented by means of statements and cases. The participants were asked to indicate how much s/he agreed with each statement.
3. Working with the WebCT and CLC portals: This section contained a number of questions for which the students needed to work, explore and analyze the two e-learning portals separately. The questions about the two portals were not completely the same because the students were already familiar with the WebCT portal, while they were not familiar with the CLC portal. Therefore, for the CLC Website we also included questions related to the user's expectations.

The participants were asked to give their general impression, and to give their opinion about the images and the icons that were showed in the portal.

We also have asked the students to answer questions while performing some tasks, like: "What does this picture/icon/key shows?" and "What kind of information do you think you will get when you click on this object?". Afterwards, the student was asked to click on the object and to look to the information that was displayed. Then, he had to answer the following questions: "What do you think of it? Is this what you expected? If not, how is it different?". At the end of the task, s/he was asked whether some other pictures or icons would have been better to represent the information.

For the CLC portal, we started by asking questions that would allows us checking if the student's expectations based on his/her first impression was corresponding with the actual intention of the Website. E.g., "What kind of information do you think this Website is offering?", "Who do you think is the target audience of this Website?", and "Which items on the screen make you think this?".

After the students finished their questionnaire, we asked them in person a few questions on what he/she thought about the portals and the culture values.

[1] http://elearning.iugaza.edu/
[2] http://www.webct.com
[3] http://www.owcp.net/clc

4 Summary of Results

We first describe the outcome of the survey for each section of the questionnaire.

4.1 Participants' Characteristics

The participants of the experiment were students from Palestine and Belgium. The students from Palestine were selected from the Islamic university, [4] Al-Azhar university, [5] and Palestine open university; [6] for Belgium, students from Vrije Universiteit Brussel[7] and Katholieke Universiteit Brussel[8] were asked to participate in our experiment. The participants were second and third year students from different study domains (Economy, Law, Engineering, Public Health, etc.), all with a similar general background.

Table 1. Participants' Characteristics

	Palestinians	Belgians
Target audience	42 participants 23 (55%) male 19 (45%) female Average age: 21 years	21 participants 8 (38%) male 13 (62%) female. Average age: 20 years
Internet / computer use	All participants use Internet (100%). 41 (98%) use the computer only for Internet. 32 (76%) use Internet every day and in the weekends. 5 (12%) use Internet during the week not in the weekend.	All participants use Internet (100%). 21 (100%) use the computer only for Internet. 18 (86%) use Internet every day and in the weekends. 3 (14%) use Internet during the week not in the weekend.
Internet activities	42 (100%) e-mails and online chatting. 39 (93%) finding information for work or study. 38 (90%) making school assignments.	21 (100%) on-line chatting, e-mail, finding information for work or study, making friends, making school assignments and finding information for personal purpose. 4 (19%) have own homepages and are working on it.
Language preferences	39 (93%) Arabic 3 (7%) English.	11 (52%) Dutch & English 7 (33%) Dutch. 3 (14%) Dutch, English and French.
Cultural preferences	19 (45%) Palestinian culture 9 (22%) Arabian and Islam 8 (19%) Palestinian & Islam 6 (14%) Arabian culture	9 (43%) Belgium culture. 6 (29%) no particular culture. 3 (14%) Flanders culture. 3 (14%): USA, Arabic culture and one consider himself belonging to the German culture
Movies preferences	All the 42 participants watch foreign movies.	All the participant watch foreign movies.

Table 1 reports the most important characteristics of the participants. As can be seen in **Table 1**, all the participants were familiar with Internet and used it regularly. The main Internet activities were on-line chatting, e-mail, finding information for

[4] http://www.iugaza.edu.ps
[5] http://www.alazhar.edu.ps
[6] http://www.upi.ps
[7] http://www.vub.ac.be
[8] http://www.kubrussel.ac.be/

work or study, making friends, making school assignments, and finding information for personal purpose.

Palestinian participants felt that their culture background belongs to Palestinian, Arabian and Islamic culture. One of the participants first wrote Arabian culture, then erased it and wrote "Islamic culture". From this it is clear that the student felt the need to distinguish between Islamic, Arabic and Palestinian culture. The Belgian participants indicated as cultural background Belgian, no particular culture, Flanders and some other cultures. An interesting finding was that all Palestinian participants have categorized themselves in one particular culture, while some of the Belgian participants see themselves as not belonging to a particular culture, and some felt that they were belong to a culture different from their national culture.

4.2 Cultural Evaluation

The theoretical frameworks that have been used to guide this part of study are the cultural dimensions of the anthropologists and systems designers; Nancy J. Adler[1], Edward T. Hall[8], Geert Hofstede[9], Fons Trompenaars[19], David A. Victor[20] and Quincy Wright[16]. The following cultural dimensions are used in our research study: *Human Nature Orientation[1],Individualism vs. Collectivism [8][12][17][1], Internal vs. External Control[1][17][12], Time Orientation [1], Authority Conception [18], Context[12][18][7], Gender Roles[8], Power Distance [8][16], Uncertainty Avoidance[8], Universalism vs. Particularism [17], Achievement vs. Ascription [17], Affective vs. Neutral [17], Specific vs. Diffuse[17], Experience of Technology [18], Face-Saving [18][12], and International Trade and Communication [14].*

To evaluate the impact of culture on the values associated with these cultural dimensions, the 16 cultural dimensions were presented by means of statements and cases. The participants were asked to indicate how much s/he agreed with each statement. The responses to these questions reflect how the participant's values were influenced by his/her culture. Students were asked to rate the statements from 1 to 5. The rating scale was as follows: 1 = strongly disagree, 2 = hardly disagree, 3 = agree to some extent, 4 = clearly agree and 5 = strongly agree. We have applied a chi square test to the set of data obtained. The following table (**Table 2**) reports the differences between the two groups.

As shown in **Table 2**, the cultural evaluation study shows that Palestinian and Belgian students agree on a number of cultural dimensions and share a number of cultural values. We see a little difference between Palestinian and Belgian students on the following cultural dimensions: Human Nature Orientation, Authority Conception, Context, Uncertainty Avoidance, Specific vs. Diffuse, Experience of Technology, Time Orientation, Face-Saving, and International Trade and Communication with an average differences between the two groups >0.2.

Participants' responses indicate that the dimensions Individualism vs. Collectivism, and Power Distance are influenced by their culture. The difference between Palestinians and Belgians are quite perceptible with an average differences between the two groups >0.9.

Table 2. Results of the evaluation of the students' cultural dimensions

Di. No.	Name	Palestine	Belgium
1	Human Nature Orientation.	4.1	4.3
2	Individualism vs. Collectivism.	2.1	3
3	Internal vs. External Control.	2.1	3.6
4	Time Orientation.	4.2	3.9
5	Authority Conception.	2.3	2.1
6	Context.	4.7	4.8
7	Gender Roles.	4.2	2.3
8	Power Distance.	3.6	2.8
9	Uncertainty Avoidance.	3.1	3
10	Universalism vs. Particularism.	1.8	2.9
11	Achievement vs. Ascription.	3.2	4.4
12	Affective vs. Neutral.	2.7	3.8
13	Specific vs. Diffuse.	2.5	2.6
14	Experience of Technology.	2.8	3
15	Face-Saving.	1.7	1.5
16	International Trade and Communication.	4.6	4.6

Hofstede's studies have categorized the Belgians as Individualism, while the Arab world concern for the collective rather than the individual. During the study we also have evaluated the participants' behavior next to their opinions about the cases and statements, by means of their responses when examining images, texts, news flashes. From the outcome it is clear that the Belgians are individualists but not as strong as the anthropological studies categorized them. They like to have chats with other people via Internet; they try to show their nation; they want to help people, but at the same time every one is looking after him self and is focused on his personal development. The Palestinian more show their religion, they focus on their traditions, but they are also trying to be independent and every one is looking after him self.

Regarding the Power Distance dimension, the study shows that Palestinians focus on roles, and leaders has a lot of authority and control. Hierarchy is very important, however the Palestinian participants want to change that situation but they don't believe they can. They said that if they would be in the position of the decision makers they would act as leaders, they believe that social roles are important in any kind of environment.

Between the two groups, the most distinct differences in cultural values orientation are found in the dimensions Internal vs. External Control, Gender Roles, Achievement vs. Ascription, Affective vs. Neutral, and Universalism vs. Particularism. The variation differences are >1.1.

The study found that the Belgian participants are more internalist than the Palestinians. They were upset when the computer did something they didn't expect and they blame the designer if they didn't like the Website. We found the Belgians highly universalistic. The Belgians insist on follow the rules whenever possible, favor equality between all members of the society and between the elements of the Website. For the Palestinians, relationships are more important than rules. Moreover, the Belgium society seems to be more Neutral; people control their relationships and make a distinction between relationship and friendship. They hide their feelings most

of the time. The Palestinians are a more Affective society; they express their feelings directly and relationships are more important than rules.

The cultural evaluation also showed that the Palestinian students are more religious compared to the Belgium students. All the Palestinians participants had respect for religion, and every Palestinian has a religion. On the other hand, from the 21 Belgium participants, 14 (67%) were without religion. We also noted from the remarks that some advertisements have offended Palestinian students.

4.3 Working with the WebCt and CLC Websites

This section highlight the most important finding from the practical part of the experiment. This part measures the understanding of the two portals aiming to know the influence of culture in understanding the portals. The 63 participants – 42 Palestinian and 21 Belgian- were asked to explore the portals using scenarios, to answer questions and to give their opinions about their understanding and expectations of icons and text appearing in the portal. The main results are as follow.

Findings from this part show that Belgium students think and work more practical than Palestinian (ex. Belgians are eager to discover new things, they reflect about the different elements in the Website and try to evaluate the Website, they made remarks like "Why is this site contains error programming?" or "The alignments in the website not in a good shape.". Overall, both Palestinian and Belgian students understood the goal of the portals.

During the study, we measured the understanding of the objects and icons by asking the participants to look to an icon/object and to name it, then to write down which information they expect to obtain when they would click on it, then we ask them to click on the icon/object and compare the result with their expectation. Members of both groups expressed equal acceptance and difficulties of understanding the icons and text. The following table (**Table 3**) highlights some examples.

As shown in **Table 3**, the graphical pictures did not always helped to understand what could be found on the target page. For example, by looking at the icon "Links" only 19% of the students from Palestine had some idea of what to expect, but when they actual clicked on the icon no one found what they expected.

Participants' responses indicated that when metaphors were used, these were not always correctly interpreted. The participants usually could recognize the icon as a familiar real world object, but then associated its real world use with it. For example, for the icon representing "Discussion", the Belgian participants understood the icon as "Advertisement page for the school", "Notes", and " Proposed things" because this kind of nails are used at the schools and universities in Belgium to put posters and advertisements at boards. Therefore, 62% thought that they had understood the icon, but for only 14% found what they expected.

The icon representing the "Homepages" was completely misunderstood. The target page contained the links to the home pages of the students who were enrolled in the same course. Unfortunately, none of the Palestinian or Belgian participants got the information expected after clicking on the icon. Palestinian participants expected the following: "Contact the teacher", "Discussion", "Women with a book", and "Guide women". The Belgian participants expected things like: "Course syllabus", "The materials needed for the course", "Class notes", "Students discussions" and "Information about the course".

Table 3. Comparison of Understanding measurement between Palestinian and Belgian

Meaning	Icon	Percent of students that reported some understanding of the icon before visiting the target page [Palestinian]	Percent of students for which the expectation was matching with the true meaning [Palestine]	Percent of students that reported some understanding of the icon before visiting the target page [Belgium]	Percent of students that could understand the icon before visiting the target page [Belgium]
Calendar		86%	41%	81%	62%
Chat		93%	89%	90.5%	92%
Syllabus		62%	2%	52%	0%
Links		23%	0%	43%	0%
Mail		79%	37%	100%	32%
Discussion		19%	23%	62%	14%
Homepages		77%	0%	84%	0%

We have also asked the participants' opinion about the use of these icons and have asked for better alternatives. Of the 42 Palestinian participants, 29 (69%) participants prefer to have better icons, while all 21 Belgian participants (100%) want to change the icons.

Palestinian participants also opened a discussion about "girl pictures" because some of them prefer not to have pictures of girls. Therefore, we asked them to vote if they would agree to replace these pictures with some thing else. Of the 42 participants, 17 (41%) have voted to change the pictures, 11 (26%) have voted to not change the pictures, and 14 (33%) didn't care. This response is inline with the cultural and religion values of the Palestinian society. Furthermore, all the Palestinian participants can accept such pictures if they appear in foreign Website but not in the portal of their own university. We also asked the Belgium participants if they had any objection to pictures of the girls. There were no objections at all, which is again inline with the cultural values associated with the Belgian society.

In the light of evaluating the understanding of the text, 22 (52%) Palestinian and 13 (62%) Belgian participants agreed that the text was easy to read and understandable. 24 (57%) Palestinian participants preferred to have the local language beside the English; while only 8 (38%) Belgian participants wanted to have the local language beside the English. The ones in favor said that the local language helps them to understand the Website.

5 Discussion and Conclusion

The findings from this exploratory study indicate that there is a convergence in some cultural values between students of both countries. Thanks to modern communication, Internet and multi-media, students are changing and the cultural gap between the two groups seems to decrease. However, there are still differences in some of cultural dimensions, sometimes only small differences but sometimes the differences are quite clear, like for Individualism vs. Collectivism, Power Distance, Internal vs. External Control, Gender Roles, Achievement vs. Ascription, Affective vs. Neutral, and Universalism vs. Particularism.

For students, Internet is becoming a part of their daily life. They use it to discover knowledge and their identity is influenced by the new recourses found on the Internet e.g., e-book, e-television, e-press, e-cinema. More and more, technology is incorporated into their lives. Therefore, they also expect to find it in their education. The second generation of the Web [10], including social networking sites, Wiki's, and communication tools, will further decrease the cultural gap between the nations especially for the generation using Internet. The Web itself is transformed from a so-called "Read-only Web" to a "Read-Write Web" [6], in which content is created, shared, remixed, repurposed, and passed along. Therefore, many people will participate in such a scenario and all of them will interact with the information. Therefore, a new culture will appear. This culture will be shared between all Internet users. The same holds for students who are using education portals.

This shows that Internet can influence change in cultural perceptions. However, the cultural identity does not change, most changes we notice, took place in understanding of and interaction with Website elements. These changes may result into new cultural dimensions specific for Internet users and e-Learning systems.

In the past, Hofstede's studies have categorized the Belgian society as Individualism with high Uncertainty Avoidance while the Arab world was categorized as Collectivisms. Our results show (at least for students) that the gap between the two groups is disappearing. Nowadays, those students that are using Internet and e-learning portals shared new cultural values. Those "digital natives" grow up with multi-media, learn and play in new ways, absorb information quickly, and have friends all around the world communicating with them using new media. Therefore some of the cultural differences are going to disappear while others, more fundamental ones, will be kept.

The outcome from measuring understanding of the objects and icons is for both groups rather similar. When the use of a concept or icon is not entirely clear, each group goes back to their social environment and tries to find the meaning for the concept in the real world. As there are still big differences in the way and style of life between the different cultures, each group gives a quite different interpretation to the concepts; therefore we found different meanings for the different cultures.

It is also clear that some cultural factors are still very important in the Arabian culture and should be taken into consideration when localizing educational portal to the Arabian culture. Although, the Belgian students had little comments related to their cultural values, they may also have some cultural values they insist to keep. This is probably the case for all cultures.

From this study, it is clear that for designing e-learning portals, it is not necessary to take into account all the traditional cultural dimensions investigated by anthropologists and systems designers. But it is necessary to know the target audience and to know the culture values that should be taken into consideration for this audience. In other words, we should investigate new cultural dimensions for the different cultures around the world for Web site design purposes.

References

1. Adler, N.-J.: International dimensions of organizational behavior Cincinnati, Ohio, South-Western/Thomson Learning (2002)
2. Al-Badi, A., Mayhew, J.: Designers' Perspective of Website Usability: The Cultural Dimension. ICWI, pp. 485-494 (2004)
3. Barber, W., Badre, A.: The Merging of Culture and Usability. The 4th Conference on Human Factors and the Web, pp. 112–121 (1998)
4. Dormann, C., Chisalita, C.: Cultural Values in Web Site Design. In: Proceedings of the 11th European Conference on Cognitive Ergonomics ECCE11 (2002)
5. Evers, V., Day, D.: The Role of Culture in Interface Acceptance. In: Proceedings of Human Computer Interaction, Interact'97, pp. 260–267. Chapman & Hall, London (1997)
6. Gillmor, D.: We the Media - The Read-Write Web (accessed date: January 2007) http://www.authorama.com/we-the-media-3.html
7. Hall, T.: The hidden dimension. Anchor Books, New York (1990)
8. Hofstede, G.: Cultures and Organizations: Software of the Mind. McGraw-Hill, London (1991)
9. Instone, I., Czerwinski, M., Mountford, S.J., Nielsen, J., Tognazzini, B.: Web Interfaces Live: What's Hot, What's Not? Panel in Proceedings of ACM-CHI, pp. 103–104 (1997)
10. MacManus, R., Porter, J.: Web 2.0 for Designers (2005) (accessed date: November 2006) http://www.digital-web.com/articles/web_2_for_designers
11. Mahemoff, M., Johnston, L.: Pattern languages for usability: An investigation of alternative approaches. In: Proceedings of the Third Asia Pacific Conference on Computer Human Interaction, Shonan, Japan (1998)
12. Marcus, A., Baumgartner, V.: Visible Language, Special Issue Cultural Dimensions of Communication Design. Part 2 (2004) ISSN 0022-2224
13. Marcus, A., Gould, E.: Crosscurrents: Cultural Dimensions and Global Web User-Interface Design. ACM Interactions 2(4), 32–46 (2000)
14. Quincy, W.: The Study of International Relations. Appleton-Century-Crofts, New York (1955)
15. Ruttenbur, B., Spickler, G., Lurie, S.: eLearning: The Engine of the Knowledge Economy. Morgan Keegan & Co. (2000)
16. Stengers, H., De Troyer, O., Baetens, M., Boers, F., Mushtaha, A.: Localization of Web Sites: Is there still a need for it? In: International Workshop on Web Engineering (HyperText 2004 Conference), Santa Cruz, USA (2004)
17. Trompenaars, F.: Riding the waves of culture, Understanding Cultural Diversity in Business. Brealey London (1995)
18. Victor, A.: International Business Communication. Prentice Hall, New York (1997)

Differences in Task Descriptions in the Think Aloud Test

Lene Nielsen[1] and Sameer Chavan[2]

[1] Center for Applied ICT, Copenhagen Business School, Howitzvej 60, 2000 Frederiksberg, Denmark
[2] MSC Software, Pune, India
ln.caict@cbs.dk, Sameerhere@yahoo.com

Abstract. This paper analyzes and discusses the ways tasks are described and perceived in a remote Think Aloud (TA) usability tests session. The paper includes reports from a study and the problems encountered during a session of remote TA tests. The sessions were performed as synchronous tests, where the facilitator and observers received data and managed the evaluation in real time with a remote participant. It was done using a system with audio conferencing and remote application sharing. The analysis and discussion include both a task description perspective and a cultural difference perspective and hereby adds to existing knowledge of usability testing.

Keywords: Usability, Remote Think Aloud Test, Cultural Usability.

1 Introduction

Think Aloud (TA) is one of the protocols for usability testing where the participant is asked to think aloud or vocalize his or her thoughts, feelings, and opinions while interacting with a product. TA helps the test leaded to understand what problems the participant is facing and to question him depending on his think aloud. There has been criticism that TA does not simulate normal tasks as in "real" life, the users do not annotate each action with thinking aloud while they are doing tasks. Another critique is that in TA, the participants tend to forget the actual task and just verbalize the screen text rather then speaking what they are thinking. Another technique is Retrospective Think Aloud (RTA). With RTA, users do the tasks silently and then talk about what they did afterwards by watching a videotape of their own actions.

Remote TA tests differ in many ways from the more traditional usability tests, but the main difference is the lack of contextual presence between the test person and the test leader, whether the test is performed synchronous or asynchronous (Dray & Siegel, 2004). The focus in this paper is exclusively on task description and perception in both traditional and remote usability tests. For both test settings, the task descriptions are written or oral instructions, but when remote usability tests are performed, the task is perceived by the test person in solitude and for the test leader there are – usually and with existing technologies - no possibilities of eye contact and perception of e.g. insecurity expressed through body language during the test session.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 174–180, 2007.
© Springer-Verlag Berlin Heidelberg 2007

The literature that deals with the practical implication of the think aloud test describes the test set-up, but has limited descriptions of how the tests are to be implemented in detail. Looking through literature that introduces the traditional TA test methods to students ([2]; [4]; [6]; [7]; [9]; [10]; [11]; [12]; [13]; [16]; [17] it has only been possible to find four who include task descriptions [9], [4] [12]. In the literature of user testing there seems to be at least two different suggestions for task descriptions: One type favors descriptions of tasks and the second favors identification with a user in a specific situation. The variations of procedures is in this paper coined *the task focused procedure* and *the scenario focused procedure*.

2 Remote TA Tests

Remote usability tests are gaining importance in recent days with the advancement of technology in desktop sharing. Remote tests provide a number of advantages such as:

- Remote testing substantially reduces costs, as it eliminates costs for travel and logistic.
- Testing can be done with a diverse pool of participants who are spread across the globe and not just local users.
- Specialists, who are otherwise not available as they have to travel to a test site, are now available as they can perform the test from their own location.
- Remote TA tests the participants in their real environment rather then a artificially created lab setup. The latter can result in anxiety and does not reflect real world working conditions.
- During Remote TA everyday disturbances can be encountered like a phone call or a colleague coming for a talk.
- There have been many studies showing that remote test can find more issues than lab tests and that the task completion rate in Remote TA can be higher than lab tests as the participants are not in stressed condition.

But there are also disadvantages with the remote TA test.

- The main disadvantage is the lack of contextual presence between the test participant and the test leader.
- The test leader does not get any clues of how many people are present together with the test participants, either just observing or actively assisting the participant.
- There can be many interruptions in the test. Often the participant gets phone calls, or some of his friends interrupt, there can be emails, and background noise. In our study an Internet connection went off and the test leader and the participant were disconnected.
- Communication with the participant by the test leader is another challenge as the test leader cannot observe body language. The test leader cannot gauge if the participant is tired, frustrated, or confused. Remote testing relies totally on what the participants say and their mouse movements on the screen. Only a skilled test leader can find the above issues by observing the screen movements.
- Not all participants are vocal.

- Performance is yet another factor. If the internet speed is slow then the response time is slow. Also, the desktop sharing tool requires some installations. And the test leader cannot remotely solve desktop problems at the participant end.
- Scheduling remote test also involves a lot of communication and mails and takes time to organize. If there are differences in time zones then it may result in non favorable times for the test.
- In a lab test you can easily video tape the screen movements, voice, and facial expressions. On Remote tests it depend on the tool whether it is possible to record all.
- Security is off concern too. The test leader cannot access if the screens are being captured. Some participants ask for the material or test links in advance. This will create security issue and also imbalance the test.
- If participants are to give feedback the participants prefer to send it back through email. These responses may change if the participants do fill in the questions immediately but take days to send it back. Unlike in a lab test where participants complete this immediately after the test.

3 The Study

This paper report from a study of remote TA, the tests were conducted with participants from France, Germany, Japan, UK, and US. Participants had to call a toll-free number to join the test and login to an internet desktop sharing application. The participants were shown the task on a digital format similar to a power point presentation (PPT), where they were able to switch back and forth between the test application and the tasks. They were allowed to make a copy of the tasks in the PPT and paste them into the test application.

Since the test involved adding new features in the test application, 20 minutes of training was given. This did not involve showing how things were done in the application, but introduced the user to how data was arranged in the application and what things were possible.

After a few initial tests, there was a need to rearrange the task description in order to facilitate the users to more easily understand the tasks and spend time on the application rather then wasting time understanding the task. This resulted in that the participants did not receive the same task descriptions. The participants from US, Germany and UK received task focused descriptions while the participant from Japan and France received task descriptions closer to the scenario focused task descriptions, with an included small introduction.

Example of the task focused description:

- Open the program (name of program)

Under

- Work area (name and number of work area)
- Which is in application area (name and number of application area)
- Under domain (name of domain)
- Open and view properties for source code (name) under this program

Example of the scenario focused description:

We have started a trial for a treatment of an illness (described in length). Create a report and definition of the clinical trial report.

Five tests, one from each country, were investigated in length and observations were made on the type of questions the participants asked and the way they interpreted the tasks. In the analysis, these are compared to the way the tasks were described. Finally a brief introduction to culture is presented in order to explain some of the differences in the participants' performances.

4 Analyses

The differences in the task descriptions seemed to generate distinctively differences in procedure. When the task focused procedure is followed the participants follows the task description rigorously, but seems to have difficulties in understanding the overall ideas behind the system. This is especially noticeable when more of the participants show that it is not clear to them that they have performed the task.

The participants from US and Germany had no problems following the procedures. They did not question the tasks or demanded to understand the task. The US participant found it difficult to assess whether or not a task was completed. The German participant did clearly not know what she was doing.

The English participant did not know when the task is done. In the final interview she is asked what she found confusing in the system and answers: "I'm not really sure what is meant by (name of menu 1) and (name of menu 2). Maybe (name of menu 1) means… well I'm not sure what it means". The names of the menus are core to understand the whole system and shows how capable she is to follow procedures without understand the overall ideas behind the system.

When the scenario focused procedure is followed the participant finds it difficult to remember the task, but can more easily understand the idea behind the system. In the abovementioned example of the scenario focused task description, is included the reason behind the task "to create a report" and a subset of tasks as well as a row of other tasks follow, this seem to make the participants forget the overall goal of "creating a report". A consequence of the description is that in later tasks the participants get confused about what they are doing.

The Japanese and the German participants understood the tasks, but in later tasks, the test leader had to explain that the idea behind the tasks was to create a report. The participants forgot that, as the subsequent tasks did not include scenarios, but built on the scenario created for the first task.

The scenario focused procedure proved difficult for the participants as they had to remember more and also find the task in a lengthy text, but it gave the participants a better understanding of the overall idea behind the system. In the task focused procedure the test leader could see that the participant followed the tasks procedures, but it made the test leader loose track of whether the participant understood what he was doing. This proves an even bigger problem, due to the constraints of the remote TA, where communication is restricted and tacit observations not possible.

As mentioned earlier the remote TA involved a short introduction/training of new features. This created complexity, as users started to compare the training information and expected the same behavior in the application.

The participants also forgot to think aloud after some time and read only the text on the screen rather then speaking out their thinking process.

There was a challenge for the participants if it involved a series of sub tasks, e.g. if the task description was big with multiple sub tasks, they would forget where to start.

5 TA Tests and Culture

The issue of cultural differences has for some time been discussed within the area of HCI. Most studies of cultural differences falls within design of user interfaces [3], [19], [1]. [14] report from a literature study of cultural issues in UI design and presents different studies that incorporates or studies cultural differences in the attitude to interfaces and to usability problems. Their study reflects upon the differences in attitude, but do not reflect upon, whether or not the methods used for testing the UI favors one culture over the other. Other reports of differences in cultural aspects in usability studies contain implications of effects on the results between similarity in culture of interviewer and test person and of no similarity [18], [15], [5].

An approach outside HCI is the psychologist Nisbett, who reports from several studies that the easterners view the world as holistic while Westerners see the world in an atomistic view. These differences in viewpoint might have an impact on the way tasks should be presented, with easterners in want of a more holistic view of a task and for westerners to reject holistic descriptions, to "get to the task". [8] (pp. 109). Returning to the above mentioned instructional literature of how to perform tests, the literature does not consider cultural differences in the test set-up or the task descriptions. Another reason why the literature does not discuss cultural differences is the underlying assumption, that a test result is only valid if task descriptions are identical for all participants.

In our study we did not encounter cultural differences in the understanding of the tasks and the sample is too small to make observations of this kind. We did encounter differences between the participants that can be explained with differences in culture. The European participants were more inclined to question the task and the system and to comment on them. They did not withhold their opinions and questioned the system immediately. The German participant found that the system she was used to worked better and encouraged the test leader to take a look at it. The French participant questioned the tasks and had an overall urge to discuss as can be seen in the following: "Test leader: What is happening? Participant: I don't know, it is your system." This was quite different for the US and Japanese participants who seemed more accepting of the situation. This, Nisbett explains with differences attitude towards either dialectical approach (Easterners) or logic reasoning (Westerners) [8] p. 37. It can be argued that the European culture favors an active debate and individualistic attitude in an egalitarian culture, the latter might explain why the US participant did not have the same attitude as the Europeans.

An observation was made with some participants not present in this study. It was the same tasks, but the participants were Indian pilot participants. These participants succeed in solving the tasks more than other country participants. The reason might be that the Indian nature encourages investigation and the finding of answers. The participants did not give up and did not declare that they could not solve the task.

They do not complain either. There was a tendency to learn the system immediately, rather then complaining on about a bad design.

6 Conclusion

In the task focused procedure the participants tend to complete the task as following a user manual. They do not think of real goals and report grammatical and UI standard based issues. E.g. in our study, one of the users was use to desktop systems and was trying to do a right click action on a web item.

The scenario focused procedure seems to apply a better understanding of the context to the participants and they are innovative in finding the solution. In our study these users reported interaction issues e.g. "This set of items should be in different tab".

If the task description is at length or more technical in nature, the TA creates a distraction in the minds of test participant. This creates a need to both a break down of length and a break down into simple language too.

The Remote TA creates both advantages and difficulties given the time and remoteness of participants, the TA also involves continuous reminders from the test leader to the participants. And finally the test leader also has to make his judgments by observing the participants' screen movements.

In summary, usability test results are dependent on how the task is written, the length of the task descriptions with sub tasks and of the type of cultural mix of the participants used.

Acknowledgements. This study was co-funded by the Danish Council for Independent Research (DCIR) through its support of the Cultural Usability project.

References

1. Barber, W., Badre, A.: Culturability: The Merging of Culture and Usability. 4th Conference on Human Factors & the Web. Basking Ridge (1998)
2. Jordan, P.W.: An Introduction to Usability. Taylor & Francis, London (1999)
3. Marcus, A., Goul, E.W.: Crosscurrents: cultural dimensions and global Web user-interface design. Interactions 7(4), 32–46 (2000)
4. Molich, R.: Brugervenligt webdesign.: Ingeniøren Bøger. København (2001)
5. Murphy, J., Howard, S., Kjeldskov, J.: Playing away from home - usability testing in a global world. CSI Communications 29(3), 18–24 (2005)
6. Nielsen, J.: Designing Web Usability. New Riders, Indianapolis (2000)
7. Nielsen, J., Mack, R.L.: Usability Inspection Methods. John Wiley and Sons, New York (1994)
8. Nisbett, R.E.: The Geography of Thought. Nicholas Brealey Publishing, London (2005)
9. Preece, J., Rogers, Y., Sharp, H.: Interaction Design. John Wiley and Sons, Chichester (2002)
10. Preece, J., et al.: Human-Computer Interaction. Addison-Wesley, Harlow (1994)
11. Rose, K., Sørensen, N.: Brugervenlighed i praksis - en håndbog. Frydenlund. København (2004)

12. Rosson, M.B., Carroll, J.M.: Usability Engineering. Morgan Kaufmann Publishers, San Fransisco (2002)
13. Rubin, J.: Handbook of Usability Testing. John Wiley and Sons, New York (1994)
14. Shen, S.-T., Wooley, M., Prior, S.: Towards culture-centred design. Interacting with Computers 20, 1–33 (2006)
15. Shi, Q., Clemmensen, T.: Cultural Usability -The Effects of Culture on Usability Test. The 6th Danish Human-Computer Interaction Research Symposium. Denmark, Århus (2006)
16. Shneiderman, B., Plaisant, C.: Designing the User Interface. Pearson, Harlow (2005)
17. Snitker, T.: Breaking Through to the Other Side. Nyt Teknisk Forlag (2004)
18. Vatrapu, R., Pérez-Quiñones, M.A.: Culture and International Usability Testing: The Effects of Culture in Structured Interviews. Journal of Usability Studies 1(4), 156–170 (2006)
19. Yeo, A.: Cultural user interfaces: a silver lining in cultural diversity. ACM SIGCHI Bulletin. 28(3), 4–7 (1996)

The Use of Cognitive and Social Psychological Principles in Field Research: How It Furthers Our Understanding of User Behaviors, Needs and Motivations, and Informs the Product Design Process

Krisela Rivera and Elissa Darnell

eBay, Inc. 2145 Hamilton Avenue,
San Jose California 95125

Abstract. Field research methods (also known as Ethnography) are useful in gathering user requirements, informing product direction, and identifying user needs and barriers. This paper will focus on how we perform data analysis for the Horizontal Visits sub-area. Horizontal Visits help identify user patterns and behaviors that inform product strategies and inspire product innovations. This paper introduces how *psychological principles and deep dive analysis* are helping eBay build better products and more useful features for its customers. Specifically, this research tried to deeply understand how and why people buy products. The study investigates users' approach to buying, attitudes, mental models, and needs. We learned that after an initial analysis of the data is completed one should continue to drill down into the meaning of the data by applying Cognitive and Social Psychological Principles to help team members more deeply understand the overall behaviors and motivations behind users' actions.

Keywords: global market, psychological principles, field research, ethnography, product development, design process, user experience research.

1 Introduction

Field research methods (also known as Ethnography [3]) are useful in gathering user requirements, informing product direction, and identifying user needs and barriers. This research method involves direct, first-hand observation of participant's behaviors in their own context (e.g., home or office). At eBay, we call this form of field research the "Visits Program". Visits provide both tactical and strategic insights to inform product definition and design. The Visits program has three sub-areas which include *Verticals Visits, Horizontal Visits, and Recurring Visits*: Vertical Visits focus on a specific area of the eBay website such as Registration or Seller Tools, etc. Horizontal Visits are more holistic and deliver strategic insights to address business questions across the organization as well as those relevant to ongoing global initiatives. Recurring Visits enable eBay employees to really understand the customers and be immersed in the user experience of the eBay site.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 181–185, 2007.
© Springer-Verlag Berlin Heidelberg 2007

The Visits program is one among several user research methods such as lab testing and Surveys. It is the ethnographic research in which up two eBay employees go with a User Experience Researcher into an eBay user's home. Visits are usually conducted in customer's homes, offices, or wherever they use eBay (e.g., in Taiwan many of the users were much more mobile and tended to use eBay in transit, so many Visits occurred in local internet cafés).

This paper will focus on how we perform data analysis for the Horizontal Visits sub-area. Horizontal Visits help identify user patterns and behaviors that inform product strategies and inspire product innovations.

Over the last two years, our approach to analysis and insight generation has changed to be grounded in social and cognitive psychological principles. These changes have increased eBay's awareness and demand for insights generated from this research. This increased demand and recognition further confirms that our new approach to data analysis is beneficial and important to the product development teams throughout eBay. We believe this new approach to analysis is ultimately helping us build better more user-centered products for the eBay customer. Visits Research have heavily shaped eBay's product decisions, future product direction and business strategies.

This paper introduces how *psychological principles and deep dive analysis* are helping eBay build better products and more useful features for its customers. The core elements include field research, analysis of field data, analysis of quantitative data (survey and dating mining), grounding of analysis with principles of human behavior such as Cognitive and Social Psychological Principles. This integration of data analysis methods has provided us with a better and more comprehensive understanding of the user insights enabling us to proceed with more confidence in building truly useful and appealing products for our customers.

Quantitative research (pre and post-Visits surveys as well as datamining metrics) are used in our analysis to help us validate insights as well as effectively identify peoples' perceptions and attitudes. Combining qualitative and quantitative data has helped us both understand what people are doing and enabled us to measure how pervasive and frequent an issue or belief is in the population. Providing both data types has also helped build a more robust and complete picture and has inspired product innovations and strategies. While field research is a powerful method to elicit a deeper understanding of "*why*" and "*how*" a problem exists, quantitative research allows us to see how pervasive or important a problem is. Again, the most effective approach is a combination of multiple methods and data analysis to gather a deeper understanding.

By combining multiple different methods and data analysis techniques together, it creates for a powerful instrument for gathering user insights and measuring data which results in more meaningful and useful products for our customers.

1.1 Goals of Buyer Research

One such Horizontal Visit we recently conducted focused on how people shop both online and at brick and mortar stores. Specifically, this research tried to deeply understand how and why people buy products. The study investigates users' approach to buying, attitudes, mental models, and needs. Another goal of this research was to

understand what drivers influence selection of a preferred shopping destination, with the purpose of identifying ways to improve the eBay Marketplace. Following are examples of some of the questions explored in this study: how do events or circumstances affect buying patterns, what motivates people to buy on eBay, and what barriers and pain points are faced by our buyers.

A combination of qualitative and quantitative techniques were utilized. Data Mining was used to explore participants' prior buying behaviors. Examples of variables examined were: time of buyer registration, time between buyer registration and first purchase, number of items purchased per month, meta-categories of purchased items, feedback score, median price per item and price range for all items, average gross merchandise bought (GMB), etc.

A post-Visits survey was also employed to further explore the research hypothesis generated while on the field visits. Methods used during the field visits included a free association task, observation of user's purchase behaviors on and off eBay, structured and unstructured interviews as well as a walkthrough of their home where they showed us everything they had purchased in the last 12 months both on and off eBay.

Each session was conducted with one participant and 1 User Experience Research accompanied by 2 eBay employees. The duration of each session was approximately 3 hours. Typically there were 2 sessions each day over a 2 month period. Visits took place in customer homes, offices, or wherever they commonly used eBay (e.g., Internet Café or Library). Each participant was compensated $50.00 for their time and an additional $50.00 for purchasing an item of their choosing from their favorite shopping website and $50.00 for purchasing an item on eBay.

2 Data Analysis and Synthesis

Both written notes and video tapes were used to capture user comments and data during the session. Immediately following each session the User Experience Researcher and accompanying eBay employees met at local café to debrief on their learning and insights. Team members discussed information such as environmental cues and influences, family situations, key behavioral insights, interesting work processes and tasks, as well as barriers, pain points, work-arounds, and user needs. After discussing all session details team members would then discuss big picture insights, motivations, drivers, and initial thoughts on product recommendations to meet customer needs. This type of data synthesis and debriefing was performed for each customer immediately following the session and lasted about 1 to 1 ½ hours.

Forty-five visits were conducted concurrently by six User Experience Researchers across four countries – Paris, France, Shanghai, China, Vienna, Austria, and the San Francisco Bay Area in the United States. After all the Visits were completed, the User Experience Researchers re-grouped to synthesize the data. To facilitate the data analysis and synthesis, several storyboard sessions were run to share insights utilizing pictures and short summaries. Team members identified, discussed and recorded holistic and generalizable trends.

The findings were then classified into high-level findings and key insights. The findings were than re-analyzed by pairing them with social and cognitive psychological principles seen in real world behaviors. This allowed us to share not only the

insights of how users' were behaving and operating but it let us generate a deeper understanding of why users behave the way they do. We found that these key behaviors were predictive regardless of shopping sites used. These behaviors were than illustrated via video clips showing how regardless of website used key behaviors could be a predictor because they were grounded in how people process information universally.

3 Insight Generation and Deliverables

Cognitive and social psychological principles apply globally. Habitual behaviors [2] and Memory and Recognition [1] are just a couple of these cognitive and social psychological principles. The principle of "Habitual Behavior" says that people are "creatures of habit" – when driving they tend to follow the same route day after day, sit in the same place for dinner every night. This principle applies to website behavior as well. We have found that people don't tend to go across the breadth of a web site but instead develop patterns of use. This pattern of use is seen across most websites. The cognitive psychology principle about recognition and recall memory shows us that people are much better at recognizing information than recalling it [1]. As applied to websites, we have found that websites that rely more on recognition by reminding people of what they were looking for previously are much more effective then those that require users to recall their previous searches. Websites that leverage recognition memory are much better at encouraging impulse buying and repeat visits than those that don't. Without memory aids people have more difficulty figuring out what to buy. These are just a few examples of the many cognitive and social psychological principles we can apply. Usability issues are also found globally but vary by degree based on users experience with technology/computers rather than the specific country in which they reside. However, unique local differences still need to be accounted for.

Our deliverables from this research has taken several forms - presentations covering not only the insights but also how cognitive and social psychological principles played an important role in understanding human behaviors and design needs. The presentations were done in PowerPoint with accompanying video clips and large print posters hung around the room to give meaning to the big picture and holistic insights. In addition, we developed a number of individual reports that addressed key questions for specific teams across eBay. These reports included the top 10 insights by area as well as overall key insights that span across the site. Today, several new products and initiatives are being developed and worked on because of this research. Unfortunately, because this work is still in progress, we cannot share the details at this time.

4 Conclusion

We learned that after an initial analysis of the data is completed one should continue to drill down into the meaning of the data by applying Cognitive and Social Psychological Principles to help team members more deeply understand the overall behaviors and motivations behind users' actions. This has helped design teams to ground their

designs in principles that can be generalized across users regardless of country of residence. We have found that some key behaviors elicited by users such as "how they process information" appear to be universal.

In addition to the behavioral insights we gleaned, we have learned that if something is difficult to use in one country it is usually difficult to use in another country. However, the degree of difficulty tends to vary based more on user expertise levels and knowledge of eBay and computers rather than country differences. Having said this, we do still find local differences in terms of attitudes, perceptions and user needs. These differences tend to be rooted in environmental differences and attitudes. For example, we find that German users prefer bank transfers over credit cards, due to differences in societies' technical and social infrastructure.

Furthermore, we have found that people have different expectations about technology, but these expectations appear to be grounded in life experiences and learnings. Finally, we see differences in perceptions and attitudes across country sites. For example, Asian users prefer animation, certain colors (e.g., pinks), cute symbols and figures while users in the US and Europe prefer straight-forward text and graphics. Visual appeal also varies based on domain (e.g., category of item). For example, people expect bank institutions to be clean organized and professional where as they expect Disneyland to be fun, cute and lively.

In conclusion, we have learned that we can design based on key Cognitive and Social Psychological Principles and that these principles apply globally. Although, localized unique features are still needed to meet the varying technological and environmental differences.

Acknowledgements. I want to thank the eBay User Experience Researchers Kaari Baluja, John Cheng, Maureen Fan, Michael Morgan and Jeralyn Reese for helping to conduct this research. I also wish to give special thanks to Larry Hannigan and Barbara Isa for the dedication and persistence in recruiting eBay customers and scheduling eBay employees. I would also like to thank my manager Ken Farmer and the Director of the User Experience Research group, Christian Rohrer, for their constant support and encouragement.

References

1. Anderson, J.R.: Cognitive Psychology and Its Implications, 4th edn. W.H. Freeman and Company, New York (1995)
2. Cialdini, R.B.: Influence Science and Practice, 4th edn. Allyn & Bacon, Needham Heights, MA, A Pearson Education Company (2001)
3. Hammersley, M., Atkinson, P.: Ethnography, 2nd edn. Routledge Taylor & Francis Group, London and New York (1995)

The Role of Annotation in Intercultural Communication

Tomohiro Shigenobu[1], Kunikazu Fujii[2], and Takashi Yoshino[3]

[1] Language Grid Project, National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan
shigenobu@nict.go.jp
[2] Graduate School of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama, Japan
s075065@sys.wakayama-u.ac.jp
[3] Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama, Japan
yoshino@sys.wakayama-u.ac.jp

Abstract. In intercultural communication, there are large barriers when the languages and the cultures are different. It is undoubtedly preferable for people to have smooth communications using their mother language. Therefore, we have developed a chat system called AnnoChat. AnnoChat has an annotation function for smooth intercultural communications. We applied AnnoChat in experiments with Japanese, Chinese, and Korean speakers. The results of the experiments showed that about 70% of the added annotations were reusable as intercultural knowledge information. About 20% of the added annotations were used to supplement information that could not be described while chatting. It is thought to be an effective example of applying annotation in intercultural communications.

Keywords: Intercultural Communication, Machine Translation, Annotation, Computer-Mediated Communication.

1 Introduction

Opportunities for intercultural communication are increasing due to the spread of the Internet. The number of Internet users in Southeast Asia keeps on increasing, and about 65% of the users are non-English speakers[1]. Mutual understanding of the partner's language is the largest barrier to intercultural communication, since the users have a variety of mother tongues.

If a pair of users speaks different mother languages, they try to communicate using a mutually understandable language such as English, but this is seldom satisfactory. Therefore, collaborative work under tends to be ineffective [1,2]. Machine translation systems are an effective solution. However, the accuracy of most machine translation systems is not sufficient, and misunderstanding due to mistranslation is common [3].

[1] Global Reach: http://global-reach.biz/globstats/

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 186–195, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Although machine translation lacks complete accuracy, imperfect translation may be acceptable if the messages can be understood. Actually, several communities are now communicating via machine translation[2] , and such communities are expected to increase [4].

We think that besides communication using a common language, a method is necessary that will allow people from different countries to communicate easily. Machine translation technology has been developed as a way to overcome language differences. With the continued advances in machine translation technology, highly accurate translations have been achieved in specific fields. If translation accuracy were excellent, mutual communication through machine translation would be possible [5]. Mail, chat, and bulletin board systems using machine translation have been developed to support intercultural communication. AmiChat is a chat system that can translate chat messages into other languages through machine translation [6]. AmiChat has a machine translation engine that can translate two or more languages. This system can display the original message entered by the user as well as the message translated into two or more languages at the same time. TransBBS is a bulletin board system using machine translation and is utilized as a daily discussion space [7]. This tool provides translation services in Chinese, Japanese, Korean, Malay, and English. This system was used as a communication tool when researchers in Asia jointly developed software as one of the experiments at ICE2002 (Intercultural Collaboration Experiment 2002[3]).

An accurate translation result is generally not obtained in intercultural communications that handle a variety of messages. If people are not familiar with the other person's cultural background, there may be a lack of understanding about a message. Proper nouns, adjectives, etc., often contain culturally-specific meanings. To deal with such cultural problems, annotations added to culturally-specific words and phrases might help to improve mutual understanding. Various systems that aim to accumulate and share knowledge information have been developed [8,9,10]. Users create annotations that are shared as knowledge information in the systems. These annotations are often asynchronously applied to static documents. However, in multilingual communication, the effectiveness and availability of annotations have not been sufficiently examined.

We have developed a multilingual chat system with an annotation function and have applied it in intercultural communication situations. This system supports communication between people from different countries who each use only their native language. However, because machine translation cannot completely prevent mistranslations, this system has a back translation function to improve the machine translation output. In this paper, we describe the effectiveness and availability of annotations in intercultural communication.

2 Multilingual Chat Tool AnnoChat

2.1 Design Policy

To support intercultural communication, we have developed a chat system called AnnoChat that has a function to create annotation data. The design policy of this system is shown below.

[2] Enjoy Korea: http://www.enjoykorea.jp/
[3] Intercultural Collaboration Experiment: http://www.ai.soc.i.kyoto-u.ac.jp/ice/

1. Supporting function of annotating to words or phrases
In intercultural communication, people with different cultural backgrounds communicate with each other. Therefore, even if a message is translated accurately and formally, people may not have the same understanding of its meaning. Proper nouns, adjectives, etc., often contain culturally-specific meanings. If a user does not know the other person's cultural background, it may result in a lack of understanding about that person's message. We should consider the possibility that words and phrases in a given message can be understood differently. Therefore, a function to add an appropriate meaning as an annotation is necessary for intercultural communication. We think that a function to add annotations to arbitrary words and phrases of a message will increase the user's understanding of machine-translation-dependent communication.

2. Supporting function of inputting a machine-translatable message
Chat messages often contain typographical errors, omissions, or euphemistic expressions. Also, some spoken languages omit the subject of the sentence, and these languages are often used in chatting online. It is difficult to translate such messages accurately using existing machine translation technology. Users can confirm whether or not the translation has succeeded by showing the translated result in their mother language before sending the message. If the translated result is not good, the user can revise it to come up with a message that is suitable for machine translation. For example, if a Japanese person writes a message to a Korean person, this function first translates the Japanese message into Korean and then translates the result back into Japanese. This method enables the translation accuracy to be confirmed in the input language. Back translation allows a user to write a sentence that is machine-translatable using only the user's mother language [11,12].

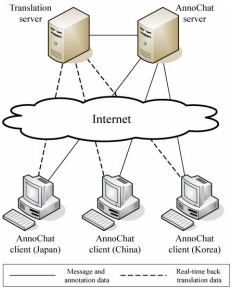


Fig. 1. System configuration of AnnoChat

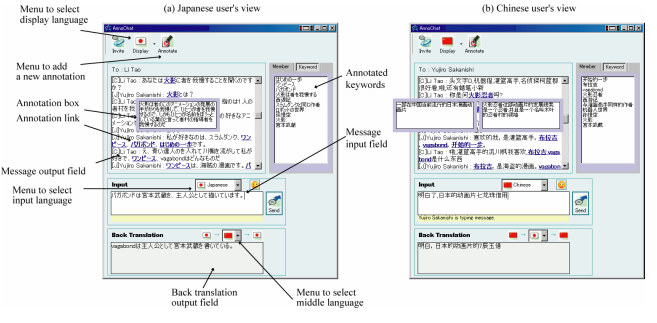


Fig. 2. An example screen of AnnoChat client

2.2 System Configuration

This system is a client-server model; the server consists of an AnnoChat server and a machine translation server (Fig. 1). The AnnoChat server receives data of messages and annotations from the AnnoChat client, and the data is translated into the other languages by machine translation. Then, the server sends the translated data to all participants of the same session. The data of the real-time back translation is communicated directly between the AnnoChat client and the translation server. The AnnoChat

server stores the message and the annotation as log data. We have used J-Server[4], which was developed by Kodensha, as a machine translation server. J-Server is available to translate between Japanese and Chinese, Japanese and Korean, and Japanese and English. We executed multi-hop translation through Japanese for languages that could not be translated directly (e.g., between Chinese and Korean).

2.3 Functions of AnnoChat

AnnoChat has a multilingual input and display, a real-time back translation function, and an annotation function. The operation procedure and screen layout are like that of an instant messenger. A sample screen of an AnnoChat client is shown in Fig. 2. Details of the functions are below.

1. Multilingual input and display

When the button to select the display language is pushed, the available language list is displayed as a menu. Displayed messages, the annotated keyword list, and the content of annotations are switched to the selected language when a user selects an arbitrary language. If a message input field is empty, selection of the input language is also switched to the selected language. Fig. 2 depicts a screenshot from a chat session between a Japanese user (a) and a Chinese user (b).

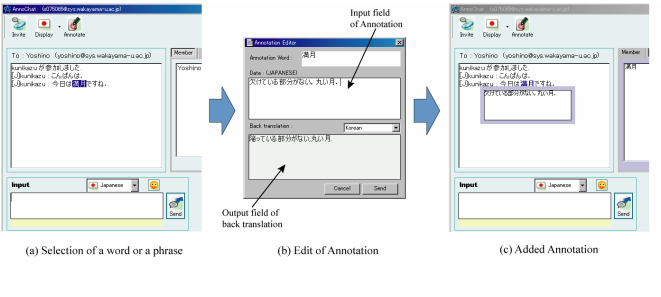


Fig. 3. Procedure for editing Annotation

2. Real-time back translation function

The client executes the back translation by using the message entered by the user at intervals of a few seconds. The result is displayed in the back translation output field. A user corrects the original message while confirming the result of the back translation displayed in the user's mother language.

3. Annotation function

The underlined bold text in the message output field shows the existence of an annotation, called an Annotation link. When the user scrolls over the Annotation link with the mouse cursor, the content of the annotation is shown in the current display language in the Annotation box. Additionally, the annotation edit window is displayed by clicking the Annotation box, and a user can edit the annotation content. The procedure of creating a new annotation is shown below (Fig. 3).

[4] KODENSHA http://www.kodensha.jp/

(a) The user selects a word or phrase that he/she wants to create an annotation for and clicks the button to create a new annotation.
(b) The user enters a detailed explanation of a word or a phrase in the annotation edit window. The user can repair a sentence by referring to the back translation result.
(c) The user clicks the send button, and the annotation data is delivered to all users.

One word or phrase often has two or more meanings. Thus, we thought that it was better to be able to create two or more annotations if necessary. This system makes it possible to create some annotations for arbitrary single words or single phrases in a chat message. Additionally, the annotation function in this system displays the annotation links to all words and phrases that appear when chatting. Words, phrases, and the content of annotations created with the AnnoChat client are translated into each language and delivered to all participants through the server.

3 Experiments and Results

3.1 Experimental Procedure

We applied the developed system in intercultural communication and examined the annotations created at the initiative of participants. The experiment was carried out as part of Intercultural Collaboration Experiment 2005 (ICE2005[5]) that research institutes from five Asian countries (China, Korea, Thailand, Malaysia, and Japan) jointly hosted in 2005. Participants in this experiment were Japanese, Korean, and Chinese. In each experiment, the participants chatted with a partner one-on-one. Chat communication experiments were carried out with 19 pairs: three pairs were Japanese and Korean; six pairs were Chinese and Japanese; and ten pairs were Korean and Chinese. The number of experiments was different with each combination because the number of participants at ICE2005 varied in each country. Participants were undergraduate and graduate students, and they had no personal acquaintance with each other. The experimental task was as follows.

1. Participants decided the chat theme mutually.
2. Participants mutually communicated a culture concerning the theme for 20 minutes using their native language.
3. After the chat task, participants created five annotations for their own messages.
4. After the experiments, participants answered a questionnaire.

In this experiment, participants created annotations after 20 minutes of chatting. Normally, it is more natural to create annotations in parallel with the communication. However, to create the annotations as a controlled experiment, we thought they should be done after some messages had been accumulated. We prepared simple themes for chatting that did not need any domain knowledge about a specific field. The participants chose one from among the six prepared themes (popular destinations for tourists, introduction of local specialty foods, etc.). We prepared the questionnaire form in English and Japanese. The participants wrote their answers in their mother tongue or in English in the description column.

[5] http://ice.kuis.kyoto-u.ac.jp/ice/ice2005.htm

In this experiment, we installed the translation server and the AnnoChat server at Wakayama University in Japan. Each client accessed both servers through the Internet.

Table 1. Classification of annotations

Classification type	The number per experiment			Ratio (%)
	Japanese/ Korean	Japanese/ Chinese	Korean/ Chinese	
Dictionary	8.3	8.2	6.5	73.5
Conversation supplementation	1.7	1.0	2.0	16.4
Translation confirmation	0.0	0.7	1.5	10.1

3.2 Results of Annotation

We examined the details of the annotations by using the annotation data created in the experiments, and the questionnaire results. Table 1 shows classified results. The investigated results are shown below.

(1) Classification of annotations

We classified the annotations created by the experiments into various usage types based on the meanings of words or phrases, the content of the annotations, and the questionnaire results. These different usage types can be used like a dictionary to explain aspects of intercultural communications. Table 2 shows examples of annotation data in experiments.

(a) Dictionary type
The participant can understand the meaning even when the annotation is added to the same words and phrases in other communications.

(b) Conversation supplementation type
The participant cannot understand the meaning even if the annotation is added to the same words and phrases in other communications. This annotation was created to supplement the content of the chat. In other words, this is dependent on the context.

(c) Translation confirmation type
This annotation does not actually explain the content of words and phrases. The annotation is created for the reason "I have not understood the meaning of words and phrases." The participant was asking the other party the meaning of words and phrases because he/she was not able to understand the result of the translation.

(2) Ratio of concordance between created annotations and requested annotations

The participants created five annotations in each experiment. The criterion for selecting them was that the user felt they were words or phrases that the partner would not understand. We investigated the difference between the created annotations and the requested annotations. The method we used to evaluate the requested annotations was a questionnaire survey after the experiments. The participants selected five words and

phrases that needed annotations in order to understand their partner's message accurately. The selected words and phrases were the requested annotations. Table 3 shows the relationship between the created annotations and the requested annotations.

Table 2. Classification example of annotations

Type	Word or phrase	Content
Dictionary	Nami Island	This site is place-name that had filmed representative drama "Winter Sonata" in Korea.
	One Piece	This is a comic and features pirates.
Conversation supplementation	I watch TV on a mobile phone.	A new service of mobile communication service carriers. Users can watch TV programs on mobile phone.
	Man district	Comic book. (* Participant corrected a typographical error.)
Translation confirmation	World edition	I can't understand this meaning of a phrase.
	Listen	This meaning of word is "listening," but this is not right contextually.

Table 3. Ratio of concordance between the created annotations and the requested annotations

	Japanese/ Korean	Japanese/ Chinese	Korean/ Chinese	Total
Number of created annotations	30	59	100	189
Number of requested annotations	30	60	100	190
Number of concordance	12	21	30	63
Ratio of concordance (%)	40.0	35.6	30.0	33.3

4 Discussion

4.1 Effectiveness of Annotations

In the questionnaire after the experiments, we questioned the participants on whether the annotations were useful for mutual understanding of the message. Table 4 shows the results of the questionnaire. Participants evaluated each item on the questionnaire on a 5-point scale (1: Strongly disagree, 2: Disagree, 3: Neutral, 4: Agree, 5: Strongly agree). Each value shows an average rating. The results are given for each pair combination (Japanese/Korean, Japanese/Chinese, and Korean/Chinese). Table 5 shows the participants' comments about the annotation function.

The participants evaluated annotations highly that were effective in helping them understand the message (Japanese/Korean pair: 3.8; Japanese/Chinese pair: 4.5; Chinese/Korean pair: 3.7). In the participants' impressions of the annotation function, a lot of affirmative comments such as, "The annotation was useful for understanding the message" were indicated. However, some participants expressed the opinion "Because the problem is solved by the conversation, the annotation is unnecessary."

Table 4. Results of questionnaire survey

Questionnaire item	Japanese/ Korean	Japanese/ Chinese	Korean/ Chinese
(1) I could communicate smoothly with a partner.	4.0	3.3	2.5
(2) I could understand a partner's message through the machine translation like own mother language.	3.8	2.8	2.0
(3) I think that annotation is useful to understand a message mutually.	3.8	4.5	3.7
(4) I think that annotating task is difficult.	2.0	2.8	2.3

Table 5. Comments about annotation function

- I want to create annotation while communicating without creating annotation after chat. (Japanese)
- When the annotation was added to proper noun of partner's country, I can understand the unknown thing easily. (Japanese)
- The annotation is useful. It is necessary for various users' communications. (Chinese)
- I thought that I could use the system easily. I can understand the content well by using the annotation. (Chinese)
- I can explain in the conversation. Should I explain by using annotation? (Korean)
- Because the translation quality was bad, annotations were useless. (Korean)

We expect that many trifling questions are generated in intercultural communication. A lot of questions in a conversation hinder smooth communications.

The classification result of the annotations showed that 16% of the created annotations were the conversation supplementation type for supplementing the content of communication (Table 1). We think that the users use the annotation function when necessary to help make the communications smooth.

4.2 Details of Annotation

In this experiment, we entrusted the object of the annotations to the participants' judgment. The evaluation value of the difficulty of the annotation-making process was less than three points (Table 4). In other words, the participants did not have the impression of difficulty. In the annotation classification, about 70% of the given annotations were the "dictionary type" that explain certain words and phrases as in a dictionary. Annotations of this type can be shared by accumulating them in the server. For instance, when the same words and phrases are input in another chat, the system can automatically display the annotation. We believe that these are useful annotations that are reusable as knowledge information.

About 20% of the created annotations were the "conversation supplementation type." Because some individual participants had felt that explanations given in the conversation itself were insufficient, they had to supplement the explanation. In other words, the annotation was used as a "second chat channel." We think that it is natural to make such chat supplementations while chatting. This result shows that the annotation function is effective in multilingual communications.

In this experiment, we entrusted the object of the annotations to the participants' judgment. The ratio of which the created annotations and requested annotations correspond was about 30% (Table 3). In intercultural communication, we confirmed that there was a big difference in the annotations that are mutually demanded. Thus, a new function to demand the addition of the annotation is necessary.

4.3 Communication Through Translation

We conducted a questionnaire survey to examine whether or not participants were able to communicate smoothly. Message understanding was evaluated highly for machine-translated messages between the Japanese and Korean languages (Table 4-(1), (2)). In the evaluation of machine translation accuracy at ICE2002 [7], a Japanese and Korean pair received the best evaluation. The evaluation of messages translated to and from Japanese and Chinese was also good. The translation accuracy of the Chinese/Korean messages was the worst. This is because the translations between Chinese and Korean used multi-hop translation through Japanese, which reduced the translation accuracy. The machine translation engine used in this experiment was the same as that used at ICE2002. Essentially, smooth communication depends on the accuracy of the machine translation. Therefore, communication supported by annotations is important because annotations enhance the translation accuracy.

5 Conclusion

In this paper, we described the multilingual chat tool called AnnoChat, which was developed for intercultural communication using machine translation. We have experimented in adding annotations to words and phrases communicated in messages in Chinese, Korean, and Japanese. In the classification of created annotations, about 70% of all annotations were the "dictionary type" that explain certain words and phrases as in a dictionary. This type can be used like a dictionary to explain unfamiliar phrases in intercultural communication. About 20% of all annotations were used to supplement information that could not be described while chatting.

We have confirmed the possibility of which the annotation makes smooth communications from participants' impression. In the future, we will support multilingual communications by improving the usability of the annotation.

References

1. Tung, L.L., Quaddus, M.A.: Cultural differences explaining the differences in results in GSS: implications for the next decade. Decision Support Systems 33(2), 177–199 (2002)
2. Takano, Y., Noda, A.: A temporary decline of thinking ability during foreign language processing. Journal of Cross-Cultural Psychology 24, 445–462 (1993)
3. Yamashita, N., Ishida, T.: Automatic Prediction of Misconceptions in Multilingual Computer-Mediated Communication. International Conference on Intelligent User Interfaces (IUI-06), pp. 62–69 (2006)

4. Salvador, C., Joaquim, M., Antoni, O., Miriam, S., Imma, S., Mariona, T., Lluisa, V.: Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation. Journal of Computer Mediated Communication, 9, 1 (2003)
5. Milam, A.: Multilingual Communication in Electronic Meetings. ACM SIGGROUP, Bulletin, 23, 1 (2002)
6. Flournoy, R.S., Callison-Burch, C.: Secondary Benefits of Feedback and User Interaction in Machine Translation Tools, Workshop paper for "MT2010: Towards a Roadmap for MT" of the MT Summit VIII (2001)
7. Nomura, S., Ishida, T., Yamashita, N., Yasuoka, M., Funakoshi, K.: Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002. International Conference on Human-Computer Interaction (HCI-03) 4, 1163–1167 (2003)
8. Wojahn, P.G., Neuwirth, C.M., Bullock, B.: Effects of interfaces for annotation on communication in a collaborative task. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 456–463 (1998)
9. Weng, C., Gennari, J.H.: Asynchronous Collaborative Writing through Annotations. In: Proceedings of the 2004 ACM conference on Computer supported cooperative work, pp. 578–581 (2004)
10. Cadiz, J.J., Gupta, A., Grudin, J.: Using Web Annotations for Asynchronous Collaboration Around Documents. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 309–318 (2000)
11. Yokoyama, S., Kashioka, H., Kumano, A., Matsudaira, M., Shiokizawa, Y., Kodama, S., Ehara,T., Miyazawa, S., Murata,Y.: An Automatic Evaluatlon Method for Machine Transiation using Two way MT. In: Proceedings of the 8th MT Summit Conference (2001)
12. Frederking, R.E., Black, A.W., Brown, John Moody, R.D., Steinbrecher, E.: Field Testing the Tongues Speech-to-Speech Machine Translation System. In: Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002) (2002)

An Activity Approach to Cross-Cultural Design

Huatong Sun

Department of Writing, Grand Valley State University, Allendale, MI 49401, USA
huatongs@gmail.com

Abstract. The demanding challenges urge us to develop an effective way to address cultural issues in IT localization and design well-localized products to support complex activities in a concrete context. This article proposes an activity approach to cross-cultural design informed by key concepts and methods from activity theory, genre theory, and British cultural studies. The approach brings cross-cultural design focus from operational affordances to social affordances.

Keywords: Cross-cultural design, localization, activity, affordance.

1 The Dilemma of Culture in Cross-Cultural Design

The demanding challenges for localization urge us to develop an effective way to address cultural issues and design well-localized products to support complex activities in a concrete context. Nowadays a large amount of today's IT products are consumer-oriented information appliances which are expected to fit into the fabric of individual user's everyday life. While the local uses of IT enterprise products in organizational contexts might share similarities in work flows and organizational structures across cultures, the local uses of IT consumer products take on various cultural and social meanings in different cultural contexts.

However, accounting for culture in current localization practices presents a dilemma. On one hand, culture takes a central role in localization process, and the term of culture is a pervasive one which appears in localization literature very frequently: one could expect to encounter the word "culture" in almost every piece of localization literature, and usually more than once. Furthermore, the importance of culture has been claimed, proven, and validated in a lot of research literature and real-world cases of market failures where companies did not carefully consider local cultural issues.

On the other hand, culture has been one of the major problems constantly hurting localization practices, where the application of the culture work constantly remains with a narrow scope and on a surface level ([24]). Localization specialists focus most of their attention on delivery aspects, such as what colors will not work for an audience in a specific country and what page layout would be preferred by some ethnic cultures. Their enthusiasm for the forms of information products—the tip of the iceberg ([13])—usually results in their ignorance of the huge underwater iceberg—the broader cultural context where information products are situated, and where products are designed, produced, distributed, and consumed. This shortsightedness results in the lack of an overall vision of localization strategies in product design and a product-oriented

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 196–205, 2007.
© Springer-Verlag Berlin Heidelberg 2007

localization process separating product design from product use. Overall, these problems cause poor usability in actual use at the user's site, culturally.

It should be noted that the notion of culture in this article is primarily informed by research in anthropology and ethnomethodology, which regards *culture* as the meanings and behaviors that groups of people develop and share over time as well as the tangible manifestations of a way of life such as artifacts and values ([9]). Thus, the local culture of a technology use should be investigated in a context where the collective and the individual meet and where the implementation (instrumental aspect) and interpretation (social aspect) interact. To envision a local culture, one might want to think of walking into a friend's messy room: initially the messiness may strike you as strange, yet have traces of familiarity, and after a while you might be able to discern the structure behind the apparent messiness.

2 Common Views of Culture in Cross-Cultural Design

A review of literature in the field indicates the shortsightedness discussed in the section above is related to three incomplete views of culture that are commonly adopted to approach local cultural issues for cross-cultural information design. In these views, culture is either limitedly defined in its entity or narrowly operationalized into practices.

2.1 Empiricist View

An empiricist view tends to approach local cultures with their experiential knowledge. This view sees the messiness and tries to log, categorize, and organize the instances of the messiness into ad-hoc localization guides. Based on personal anecdotes and empirical studies of local cultural conventions, those guides usually include lengthy lists of do's and don'ts for different ethnic cultures, and elaborate on translation, coding conventions, interfaces, formats including layouts, fonts, and graphics, and other international variables ([5], [15]). However, this approach is built on an engineering approach favoring efficiency over context-sensitivity ([24]). The whole process of localization is simplified as part of the engineering cycle from the planning stage to the testing stage detached from its use context. In the pursuit of engineering and automating this process, localization professionals only need to attend to delivery and style such as translating the user interface and resizing a dialog box. The problem we see here is that localized products and services are usually not fit for use contexts. Professionals are only working on the forms of information products.

2.2 Positivist View

To address this problem, researchers suggest bringing cultural contexts into practices and research ([2], [6], [16]). They seek to understand the structure behind the messiness, and borrow popular cultural models from the fields of international communication and management such as Hofstede's for use in localization practices. Hofstede regards culture as "mental programming," and he develops a list of cultural dimensions

such as power-distance, collectivism vs. individualism, femininity vs. masculinity, and uncertainty avoidance to describe national cultures. Compared to ad-hoc localization guides, models of cultural dimensions are more structured and more research-based ([13]). They provide vocabularies and structured frameworks to compare cultural patterns across nation, which is helpful for localization design.

However, these cultural models promote a positivist view on culture which strips rich contextual data away during the formation of the formal structure. First, only the dominant cultural values in a national culture are represented in cultural models, other subcultural factors such as the individual user's gender, age, organizational affiliation, or ethnic group are ignored. As Myers and Tan ([17]) comment, these cultural dimensions based on the concept of a national culture are "overly simplistic" (p.24). In localization practices, we often see local cultures that are related to a subculture group (e.g., instant messaging is more popular in groups of teenagers than in other age groups) in a country, but these cultural models cannot help design and localization if they are obscured by a set of national culture dimensions. Second, these views of culture place concrete cultural realities into static dimensions ([17], [24]). Some researchers who employed cultural dimensions in research work noticed that those dimensions could not fully explain the complex phenomena found in the field as the messiness and complexities of the local contexts (e.g., immediate context) are often neglected while only general patterns originating from the broader social contexts are attended to ([11]).

In fact, missing the actual practice of social activities is a common problem in localization literature, as we can see from both the empiricist and positivist approach. As an example, Hoft's book *International Technical Communication* covers many aspects of internationalization and localization with "international variables," but none of them come from field studies of use activities in context. When they follow her suggestions of cultural editing (p.123), designers can only beautify buttons with local translations, though the real goal here is to support complex user activities in their local context.

2.3 Semi-contextualist View

To capture rich activities at local sites, this group ([26], [27) approaches local cultures by contextualizing the messiness into local user activities and tying the messiness with design models from fieldwork methods such as contextual design ([3]).

However, this approach is still narrow in the scope as it only contextualizes half of the process. First, though it allows localization professionals to examine the immediate context, it fails to connect the immediate context with the broader socio-cultural context. Second, those work models were developed to examine work practices in the organizational context, but not for understanding social computing practices in the individual context like mobile phones and other information appliances. Third, some guidelines about cultural issues are limited and superficial. A common limitation for current fieldwork methods is that they just focus on the aspect of tool-mediated production of an IT artifact in context, but rarely explore its sign-mediated communication. Thus they are good for gathering design requirements for instrumental convenience, but weak for exploring design options for social affordances.

3 An Activity Approach to Cross-Cultural Design

To seek better solutions to cross-cultural design and improve localization performance, I turn to activity theory, genre theory, and British cultural studies which study cultural and contextual factors from different angles forming a broader understanding of local cultural factors complementarily. Key concepts and methods from these theoretical constructs are borrowed and integrated to develop an activity approach to cross-cultural design. This approach examines user activities in local contexts. It regards usability as a mediation process consisting of an instrumental aspect (mediation of practices) and a social aspect (mediation of meanings). This new methodology has a structured and flexible framework to investigate concrete uses via fieldwork with a robust structure to attend to cultural factors in both the broad socio-cultural context and in the immediate context.

3.1 Activity Theory: Examining Concrete Use Activities in Local Contexts

As a cultural-historical approach, activity theory claims that people's activities are an object-oriented and tool-mediated process in which actions are mediated through the use of artifacts (including tools and languages) to achieve a transformative objective. It is significant for the field of HCI to explore cultural and contextual issues by bringing the following valuable concepts and principles to practice and research. First, a focus on the tool (or artifact) on the basis of activities from activity theory helps us see how a technology is interpreted as an object used by people to perform activities in context. A tool becomes a tool only through use. Therefore, a tool needs to be studied in its use setting; it is not meaningful to study a tool in isolation.

Second, all human activities involve the use of tools, and activities are mediated by tools. The concept of mediation is valuable here as it shows the ways that people use artifacts are socially, culturally, and historically determined. And the emphasis of activity theory on the mediation process, the transformational objective, and the activity system suggests a process-oriented view of the design process rather than a product-oriented view.

Third, activity theory uses an activity as the unit of analysis to study human activity and tool mediation, which brings the vision of contexts into the object of inquiry. The activity system includes "a minimal meaningful context" ([14]). In this "minimal meaningful context," history, development, meanings, community, rules, and even culture are articulated into a unified framework, which makes the context consideration an inherent feature of activity-theory-based HCI research.

Fourth, the three-level structure of activity makes it possible to distinguish and describe contextual factors as associated with the instrumental aspect or the social aspect of an activity. According to Leont'ev, the unit of activity is hierarchically structured on three functional levels: activity, action, and operation. A concrete activity is always motivated by general objectives acknowledged and recognized in the local community and in the socio-cultural context. The concrete activity is realized by actions which are goal-directed in an immediate context (e.g., at the workplace or at home). Actions are usually conscious, and they are similar to the "tasks" we often talk about. An action is realized by conditions in a use situation (i.e., a material setting). Operations are usually non-conscious and automatically performed. For example, a

concrete *activity* involves a user who wants to maintain regular contact with an old college friend by sending messages of greetings occasionally. As she does not want to disturb her friend who might be busy at that moment, she chooses text messaging for communication. The act of sending a text message to the friend is *action* here. *Operation* refers to the mundane details when the user interacts with cell phone keypad and text messaging application. In all, the three-level structure is not static but fluid depending on the use situation.

The three level of activity structure brings insights to the notion of affordance by placing it in context ([1]). Affordance describes the action possibilities posed by the artifact in use and associates the artifact with practices; however, this term is widely used in the HCI field but not clearly theorized yet ([8], [20], [21]). With an activity-based framework, Baerentsen and Trettvik assert that "[a]ffordances are not properties of objects in isolation, but of objects related to subjects in (possible) activities" (p.59). They propose that the concept of affordance should be treated as a generic concept which distinguishes affordance on the operational level with "operational affordance" (e.g., the touch and feel of a phone pad in the example above), "instrumental affordance" on the action level (e.g., communicating unobtrusively), and "need related affordance" on the activity level (e.g., staying in contact with college friends).

3.2 Genre Theory: Investigating Structuring Forces Behind Habituated Uses

Genre theory attends to textual and contextual regularities, repeated actions, and technological influences, both across texts and across practices by examining social exigencies of genres ([4]). A genre is "a collection of practices that finds its nexus in the recurrent, dynamic activities in which users engage" ([23]). Genre theory brings the following insights to the exploration of cultural and contextual factors during use.

First, the notion of genre can help us better understand the artifact in a social and historical context. In HCI research, genres don't have to be textual ones, and artifacts are broadly interpreted as genres to investigate how the connection of design and use is dynamically settled in different interface features by inquiring about rules and habits related to genres. For example, a structured layout on a German website and vibrant colors on a Brazilian website present different generic features informed by local reading habits and design preferences. By providing socially constructed interpretive conventions, genres are "affordances" here to help interpret the artifact's use in context.

Second, genre theory provides a foundation for interpreting actions from a social angle. According to Miller ([18]), genres are social actions in response to recurrent situations with social motives. Dias and his colleagues interpret a social motive as "a motive that is socially recognized and allowed for" and "that the culture acknowledges you may have and allows you to have" (p. 20). As "the culture's arrangements," genres are "means of legitimately acting on these motives." In a local setting, social motives take the form of "local purposes" (p.22). Linking genre theory to activity theory, they suggest that genres are "enactments of recognized social motives" and "activities in Letont'ev's sense" (p.25). In this sense, when we found mobile text messaging was used to conduct long conversations in one culture and to have small talk in another culture, we might want to design different interface features to support different user tasks.

Third, the rule-tool relationship embodied by genres is insightful to illustrate how uses of technologies are structured in social contexts and how cultural dimensions influence a particular IT design. Influenced by Gidden's structuration theory, Miller suggests genres are capable of reproducing social structures with their recurrent nature in situated communication ([19]). Regarding a technology as a genre can help us reveal the reciprocal relationship between a technology and the social context in which it is produced and used.

3.3 British Cultural Studies: Interpreting Local Use as Cultural Consumption

Culture is political for scholars of British cultural studies, which is "a way of living within an industrial society that encompasses all the meanings of that social experience" ([7]). They are concerned with the generation and circulation of meanings in technological societies at this postmodern stage. Its emphasis on popular culture and daily life practices helps us to understand technology use in everyday life and the influence of consumer culture on IT product design and use.

The articulation model ([22]) explores contextual factors from a discursive angle, highlighting the mediation of meanings on the social aspect of human action which activity theory does not. Here articulation as a methodology maps the context, but "not in the sense of situating a phenomenon *in a context*, but in mapping a context, mapping the very identity that brings the context into focus". Thus "identities, practices, effects generally *constitute* the very context within which they are practices, identities or effects." It is a process of creating connections between various contextual factors on the level of practices and the level of meanings. As an example of such mapping, the circuit of culture examines five key processes in a development cycle of an artifact: representation, identity, production, consumption, and regulation ([10]). In the real world, these five elements continually overlap and intertwine in complex and contingent ways.

Applying the circuit of culture to cross-cultural design can show how other elements (representation, identity, production, and regulation) interact with and contribute to the "consumption" element in the whole lifecycle. It suggests that the consumption process is not the only significant and stand-alone process we need to consider in design. The idea of cultural consumption as both a material and symbolic activity directs our attention to the signifying practices and "identity values" of daily technology use.

Based on the discussion above, I propose an activity approach to cross-cultural design. With a focus on the mediation of *meanings* and of *activities* in context, this approach regards usability as a diffusing feature across the activity system, incorporates cultural factors from both the *immediate context* and *socio-cultural context* into the object of inquiry, and situates culture in the dynamic interactions of the *instrumental* and *social* affordances of the technological artifact.

4 The Case of Mobile Text Messaging

The following case is drawn from a recent cross-cultural study of mobile text messaging use in American and Chinese contexts ([25]) to illustrate the activity approach. Forty-one frequent users ranging from 18 to 30 from two sites participated in that

study. Data were collected via methods of questionnaire survey, diary study, qualitative interview, and observation. During data analysis, message diaries were studied to explore concrete use activities at the intersection of the immediate context and the sociocultural context; collected text messages were coded to examine habituated uses and search for structuring forces of this technology in a sociocultural context; and interview transcripts and observational notes were investigated to interpret the mediation of meanings and social motives in everyday life practices.

"Lily" is a 26-year-old college teacher from China. Texting is "an indispensable means of communication" in her life, and she thinks that most of her friendships and cousinship are maintained and enhanced by text messaging. Lily moved to her current city after graduation. As a stranger in the new place, she uses text messaging to stay in contact with childhood friends, college friends, colleagues, and relatives. The diary study indicates almost half of her messages were sent to exchange recent life situations with friends. She and her friend usually engaged in a conversation consisting of several message exchanges.

When I interviewed Lily, she was busy preparing for her upcoming wedding ceremony, and she had just sent out the first round of invitations for their wedding banquet to her friends via text messaging. She appreciated the affordance of getting quick feedback from text messaging. Friends typically texted her with congratulations, told her whether they would be able to come, and how many of them would make it. Especially for friends at a distance, it was more convenient to send text messages than to mail invitation cards. Lily also prepared a few paper-based invitations. These were primarily reserved out of respect for her older work colleagues with whom she had a good, yet more distant relationship.

Lily finds text messaging agrees with her personality compared to other technologies. For example, she does not like to make phone calls with friends all the time as it is abrupt to call people and ask about their recent situations after years without contact, nor does she like to go online to chat via instant messaging, as she does not feel it genuine to chat with different friends at the same time. She values simple friendships and one-to-one communication that text messaging affords. By using text messaging for maintaining and enhancing her social network, Lily actually identifies herself strongly with the socio-cultural norms surrounding her. In a collectivist culture, relationships are relatively long lasting, and individuals feel a deep personal involvement with each other. This long-term relationship orientation is mediated nicely with mobile messaging that allows people to stay in touch in an unobtrusive way. She confessed in the interview: "Sending text messages helps me understand [the saying] 'the friendship between gentlemen appears indifferent but is pure like water (Jun zi zhi jiao dan ru shui)' in a deeper way. It makes me feel good by texting and greeting friends occasionally." The phrase "the friendship between gentlemen appears indifferent but is pure like water" is a Confucian motto about how to socialize with friends. It has been told for thousands of years in China and is deeply rooted in Chinese people's daily social practices. People are taught that they should treat their friends genuinely with reserved warmth and reasonable distance. The best friendship is like pure water, maybe mild, but enduring without being tainted with personal interests or excessive contact.

While Lily enjoys the social affordance of this text messaging technology, she is also bothered by its instrumental limitations as she finds her care and consideration is

confined to the size limit of a text message. She likes to compose long and complex messages to describe life scenarios; however, "about 70% of the time" (her estimate) when composing text messages, she receives a prompt telling her that she has reached the size limit. Then, she has to go back and delete some words without ruining the clarity of her messages. It is annoying to go through this process daily, but she has no other way.

Lily's use of messaging technology indicates that her local uses were influenced by dimensional cultural factors such as high-context communication style and collectivist culture, but these dimensions are not abstract and isolated ones. They are also shaped by the local conditions in the immediate context and by Lily's own personality. In addition, gender factor and generation factor come into play. As a young female user, Lily embraces the messaging technology willingly than her older generations and tends to compose lengthy messages than male users.

With the lens of the activity approach, we can see a specific local use is developed in a concrete activity situated at the intersection of the immediate and social contexts, and this local use echoes with both the user's subjectivity and the surrounding culture's ethos. Thus, simply applying cultural conventions to cross-cultural design is ineffective; we need to develop rich understandings of use activities in context to designing local technology. Moreover, it is insufficient to tie cultural issues only to national cultural dimensions as gender/generation issues clearly affect local uses of text messaging in this case.

The activity approach brings design attention to social affordances. For example, Lily hoped that she would not have to delete words to fit in the message size limit most of the time when she texted to her old friends. On the surface level, it seems that this use problem can be easily fixed by increasing the message size limit; however, the real issue here is how to better support for longer conversations and lengthy chats and how to help maintain social network in a collectivist culture via text messaging.

It is clear that the focus of cross-cultural design needs to move from localizing for operational affordances to localizing for social affordances. Users value social affordance during their use. The translation of menus is an affordance on the operation level, but what users really want and value is an affordance on the activity level. It is shocking to see that the localization work for the mobile text messaging application on most phone models only involves the translation work of the interface for operational affordances, the interface and the functions of the technology have remained the same except for the improved inputting methods after all these years, even though mobile messaging technology is used for different purposes (e.g., small talk vs. long conversation) in different cultures. The localization work informed by the activity approach would address this need and better support local uses.

References

1. Baerentsen, K., Trettvik, J.: An Activity Theory Approach to Affordance. In: Bertelsen, O.W., Bodker, S., Kuuti, K. (eds.) Proceedings of the Second Nordic Conference on Human-Computer Interaction, pp. 51–60. ACM Press, New York (2002)
2. Barber, W., Badre, A.: Culturability: The Merging of culture and usability. In: Paper presented at the Proceedings of the Fourth Conference on Human Factors and the Web, Basking Ridge, New Jersey (1998)

3. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufman, CA (1998)
4. Dias, P., Freedman, A., Medway, P., Pare, A.: Chapter 2, Situating writing. In: Worlds apart: Acting and writing in academic and workplace contexts, pp. 17–46. Mahwah, NJ (1999)
5. Esselink, B.: A Practical guide to localization. John Benjamins Pub Co, PA (2000)
6. Faiola, A.: A visualization pilot study for hypermedia: Developing cross-cultural user profiles for new media interfaces. The Journal of Educational Multimedia and Hypermedia, vol. 11(3) (2002)
7. Fiske, J.: British Cultural Studies and Television. In: Allen, R. (ed.) Channels of Discourse, Reassembled: Televison and contemporary criticism, 2nd edn., pp. 284–326. University of North Carolina Press, Chapel Hill, NC (1987)
8. Gaver, W.: Technology Affordances. In: Proceedings of Conference on Human Factors in Computing Systems (CHI '91), pp. 79–84 (1991)
9. Geertz, C.J.: The Interpretation of Cultures. New York: Basic Books (1973)
10. Hall, S. (ed.): Representation: Cultural representations and signifying practices. Sage, London (1997)
11. Harvey, F.: National cultural differences in theory and practice: Evaluating Hofstede's national cultural framework. Information Technology & People 10(2), 132–146 (1997)
12. Hofstede, G.: Culture and Organizations: Software of the Mind. McGraw-Hill, New York, NY (1991)
13. Hoft, N.L.: International technical communication: How to export information about high technology. John Wiley & Sons, NY (1995)
14. Kuutti, K.: Activity theory as a potential framework for human-computer interaction research. In: Nardi, B. (ed.) Context and Consciousness: Activity Theory and Human-Computer Interaction, pp. 17–44. MIT Press, MA (1996)
15. LISA. The localization industry primer. Switzerland: Localisation Industry Standards Association (2003)
16. Marcus, A., Gould, E.W.: Cultural dimensions and global Web user-Interface design: What? So what? Now what? Paper presented at the 6th Conference on Human Factors and the Web, Austin, Texas (June 19, 2000)
17. Myers, M.D., Tan, F.B.: Beyond Models of National Culture in Information Systems Research. In: Tan, F.B. (ed.) Advanced Topics in Global Information Management, vol. 2, Idea Group Publishing (2003)
18. Miller, C.R.: Genre as social action. Quarterly Journal of Speech 70, 151–167 (1984)
19. Miller, C.R.: Rhetorical Community: The Cultural Basis of Genre. In: Freedman, A., Medway, P. (eds.) Genre and the New Rhetoric, pp. 67–78. Taylor & Francis, London (1994)
20. Norman, D.A.: The Design of Everyday Things. Basic Books, New York (1988)
21. Norman, D.A.: Affordance, Conventions and Design. Interactions, pp. 38–42 (May/June 1999)
22. Slack, J.: The theory and method of articulation in cultural studies. In: Morley, D., Chen, K.-H. (eds.) Stuart Hall: Critical Dialogues in Cultural Studies, pp. 112–127. Routledge, NY (1996)
23. Spinuzzi, C.: Designing for lifeworlds: Genre and activity in information systems design and evaluation. Unpublished PhD's dissertation, Iowa State University, Ames, IA (1999a)
24. Sun, H.: Why cultural contexts are missing: A rhetorical critique of localization practices. In: Proceedings of STC 49th Annual Conference, Nashville, TN (2002)

25. Sun, H.: Expanding the scope of localization: A cultural usability perspective on mobile text messaging use in American and Chinese contexts. Unpublished doctoral dissertation, Rensselaer Polytechnic Institute, Troy, NY (2004)
26. Vaananen-Vainio-Mattila, K., Ruuska, S.: Designing Mobile Phones and Communicators for Consumers' Needs at Nokia. In: Bergman (ed.) Information Appliances and Beyond: Interaction Design for Consumer Products, Academic Press, San Diego (2000)
27. Yu, L., Tng, T.H.: Culture and design for mobile phones for China. In: Katz, J.E. (ed.) Machines That Become Us: The Social Context of Personal Communication Technology, pp. 187–198. Transaction Publisher, New Brunswick, NJ (2003)

Creating an International Design Team

Becky Sundling

Microsoft China,
Mobile and Embedded Devices User Experience Team (MEDX),
Beijing, China
beckysun@microsoft.com

Abstract. The Microsoft Mobile and Embedded Devices User Experience Team (MEDX) is made up of 20 designers and researchers at the main headquarter office in Redmond, outside of Seattle, USA. In the spring of 2005, a design team of four people was started in Beijing, China. How does one successfully set up a remote team when collaboration is central to the task? What are the realities of creating clear communications across diverse languages, cultures and time zones? How does one create an appealing career path for remote talent? This paper will discuss the challenges, celebrations and lessons learned during the first two years of MEDX Beijing's development.

Keywords: User Experience, International Design Team, Interaction Design, Visual Design, Remote Team, Beijing, China, Mobile Design, Microsoft.

1 Introduction

MEDX had several goals for building a design team in Beijing:

- Leverage amazing China design talent
- Develop first-hand understanding and local expertise of the China Market
- Support China field research and gather specific observations
- Collaborate with MED product development teams in Beijing
- Build relationships with Microsoft Research Asia

The initial responsibilities of the Beijing team included creating visual design assets for Windows Mobile, participating in mobile interaction design projects, and providing China-focused perspectives and insights to Redmond. The initial team consisted of one American senior designer with previous Microsoft experience, two local Chinese visual designers, and one local Chinese interaction designer.

The long-term vision for the Beijing team is to build enough expertise and skill so it will own end-to-end development of both user research and design initiatives for China and appropriate global market products.

2 Process

Our main point of contact in Redmond was MEDX Art Director Greg Melander. We established our communication rhythm through bi-weekly conference calls in which

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 206–211, 2007.
© Springer-Verlag Berlin Heidelberg 2007

we discussed new projects, reviewed previous work and clarified questions. We relied on two very simple tools: the phone for voice (internal VOIP for cost-efficiency), and MSN Instant Messenger with a webcam.



Fig. 1. A creatively doctored MSN IM screenshot with Peter Chin (hidden) and Greg Melander in Redmond; Yin Zhu, Liang Chen, Becky Sundling, Rokey Zhang in Beijing

We quickly discovered the webcam was a critical component in our communication. We were able to point it at white boards, share sketches, or use body language to clarify a term. It also helped us keep up with new haircuts, glasses and fashion statements. The importance of this simple visual connection could not be underestimated.

It provided a face and emotional connection to the rest of our team, located on the other side of the world. As designers, words sometimes elude us, where a simple drawing or hand movement can transcend language and language barriers. Without this ability to communicate visually, many nuances would be lost.

We tried several other collaboration tools but the complexity seemingly outweighed the value.

Initially it was critical to have two main points of contact to facilitate our communication process. Greg gathered and prioritized projects, and then communicated the details. My role was to ensure clarity and to make sure the team knew specifically what was needed, in what priority, and by when. Many times during conference calls I would draw pictures to clarify a concept, act it out, or stop the conversation to backtrack and verbally explain a design term or slang phrase. I watched their reactions during the calls to make sure they understood what was necessary. I also acted as a bridge to help my team understand Microsoft culture and unspoken corporate expectations.

3 Differences…Challenges

Remote communication of people from the same country would create process challenges; add a mixed culture team new to the industry and Microsoft, and you have an interesting combination of factors. Following are some of our challenges organized into three categories: (3.1) Process and Remote Workflow, (3.2) Culture, and (3.3) Organizational Challenges.

3.1 Process and Remote Workflow

Following are some of the top issues we encountered regarding process:

- Lack of visibility of the larger vision and context to better understand projects
- Impossible to participate in spontaneous discussions and decisions
- True collaboration proved extremely difficult
- Communicating across time zones required flexibility and balance

The first two points make the third true. In a collaborative, iterative environment, it takes a lot of effort, time and dedication to keep a remote team up-to-date on the latest turn of direction. Visual design projects overseen by one individual could be communicated and then successfully delivered through several rounds of iterations. This process worked well as the ownership and approval model were very concrete. However, interaction design projects proved to be more challenging, mostly due to the exploratory nature of multiple ideas moving forward at the same time, and then reviewing each other's work to choose a direction and combine thinking. We finally settled on a model where one designer in Beijing worked on a fairly defined problem with direction from one UX Lead in Redmond. This process worked well, though the Beijing designer couldn't participate in the Redmond user tests of his designs.

With a very short overlap of convenient working hours (8am-10am Beijing is 4pm-6pm Redmond) and only four overlapping work days (Beijing Saturday is Redmond Friday) scheduling meetings can be a challenge. While some individuals will actually call into a 3am meeting, teams need to work with each other to settle on a livable

compromise. Having consistent times to talk about issues increases the quality of communication, as it is difficult to schedule spontaneous meetings with remote teams. Another challenge is realizing the need to communicate holiday schedules (Chinese work during American holidays of Thanksgiving, Christmas and New Years, while Americans work during Chinese New Year, Spring Festival and the National Holiday). It is important to recognize each culture's respective holidays have the same significance, and to plan these events into overall project plans.

3.2 Culture

The following points highlight cultural nuances discovered during this process:

- Unknown cultural rules and boundaries
    - Initially, the Beijing team was uncomfortable asking questions or interrupting a "superior"
    - The Beijing team was inexperienced in making decisions with ambiguous information; but given the time zones, this skill was highly necessary
    - Acting proactively and taking individual responsibility had not been well rewarded in the last 50 years of Chinese history
    - Western approaches are sometimes seen as aggressive
- Language and terminology charades
- Difficult to find desired level of design expertise
    - Hard design skills using tools are taught and fairly available
    - Soft skills of problem solving, team work, proactive decision making are less common

Initially some of the differences in working style were quite surprising. What is "natural" and taken for granted in a western corporate environment is not the same landscape in a Chinese office. We sat in a large enclosed office due to the visual confidentiality of our work, which created a safe haven for us to experiment with some of these issues. My goal was to build mutual trust and respect among the team, and for them to know it was safe and they were expected to question, clarify, challenge, and debate. This was a completely foreign concept, at least in the workplace, though they fully embraced the opportunity once they believed the sincerity of it.

Another area that needed attention was for the team to understand that *no action* was worse than a wrong action. Often, we finished a conference call to discover a question we hadn't asked. Initially this posed a block, and the designer did not know how to move forward. When this happened, we would look at available information, past examples, and attempted to imagine Greg's answer to the question. The goal was to predict a direction that would keep us efficient for another day, until we had a chance to receive the information we needed. After several successes using this process, the team naturally adapted this mode of operation.

3.3 Organizational Challenges

The following points highlight two management challenges:

- Cross-group ownership and double reporting fun
- Non-Chinese manager leading a Chinese team; a lot of cultural learning and translation is required

The Beijing team reported into the local China office, but had a dotted line to the Redmond team, who owned our priorities. At times this caused confusion as local teams wanted to work with the "China design team" but their request often fell outside our top initiatives. This could be explained logically, but it required constant effort to balance desiring local impact and staying true to original priority.

As a non-fluent foreigner managing a local Chinese staff, I could not fully participate in many activities. Sitting in a collaborative work space worked great for the team, but I could not interject suggestions or ideas a normal manager might have while they were debating in Mandarin (though their English improved greatly). I also was unable to keep a heightened pulse on morale through my own channels; I was dependent on key members to communicate issues I needed to know about.

In addition to interpersonal nuances, I was also challenged by the Beijing team that I didn't understand real Chinese culture well enough (I needed to take the bus more, to learn the language better, have a deeper understanding of the history, etc). I agreed and continued to struggle through my language classes.

4 Celebrations

One of the biggest successes has been the incredible growth of the Beijing team, mostly with core proficiencies such as becoming more proactive, confidence in communication, and verbalizing design decisions and opinions. Other celebrations included:

- Delivering thousands of high-quality visual assets for a world-wide product
- Moving toward more one-on-one communication with Redmond design members (Now the Beijing designers work directly with Redmond designers; Greg and I are no longer the only communication touch-points between the teams.)
- Starting to grow our team and ownership of projects
- Increasing awareness of design value in Beijing office
- Incredible cultural learning for all involved
- Taking the Beijing team on their first trip to the USA (Seeing the States for the first time through their eyes was amazing.)

5 Lessons Learned

There is no such thing as over-communication or being too clear when working together across cultures, locations and time. Many of the cues we take for granted by simply being from the same country or working in the same physical space need to be supplemented with concrete communication.

- The remote team is responsible for their own exposure

- Push out communications and request information. When you are physically not seen, you are easily forgotten. The remote team needs to take the initiative to engage with the headquarter team.

- Face-to-face visits are critical
  - After our visit to Redmond in August 2006, the development of the Beijing team skyrocketed. They were able to concretely understand and see a larger, experienced team work together in a collaborative environment. They had never been exposed to this type of working style, and better understood how the different team roles relied on each other to create a stronger whole.

- Participate in key management meetings
  - As a remote design lead, it is absolutely critical to be invited to and attend key management meetings. No one else will raise your issues or communicate important messages for your team.

- Email communications across time zones have a completely different workflow model
  - Depending on when the email is sent, a response from the remote team may be immediate or 12-24 hours away. This can cause confusion if people do not know they are asking for something at 3am local time.

- Make it as easy as possible for the HQ team
  - Talk in their time zone, do as much as possible to help them provide the information and inclusion needed.

Some projects work well being frequently passed back and forth over the ocean. However, as remote teams build up expertise and mutual respect with the headquarter team, natural evolution may need to include larger end-to-end ownership of appropriate projects. Providing autonomy within established guidelines will best leverage, grow and retain local talents.

This approach requires dedication and a longer-term vision toward developing the local team into more than an outsourcing solution. The headquarter team and their management need to *dedicate effort, time and resources to ensure a continued return on their initial investment,* and to provide opportunities for *the local team to take on additional responsibility*. As China catches up with the western approach of developing --your product here--, one would do well to incorporate and embrace a global perspective on product development, and to leverage the local expertise in place to develop for local markets.

6 Future Vision

Continuing the goal of creating a world-class Asian design team, MEDX Beijing is driving toward providing strategic impact for global and China mobile products. Leveraging qualities only a design team in Asia can bring, and with continued leadership and guidance from our HQ design team, delivering on this vision is definitely in sight. We look forward to defining this future with our partners in Redmond.

Incorporating the Cultural Dimensions into the Theoretical Framework of Website Information Architecture

Wan Abdul Rahim Wan Mohd Isa, Nor Laila Md Noor, and Shafie Mehad

Faculty of Information Technology and Quantitative Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
{wrahim2}@gmail.com, {norlaila,shafie}@tmsk.uitm.edu.my

Abstract. Information Architecture (IA) has emerged as a discipline that is concerned with the development of systematic approaches to the presentation and organization of online information. The IA discipline has commanded significant attention from professional practitioners but lacks in the theoretical perspective. In our effort to formalize the knowledge of the discipline, we report on the extension of our initial work of formalizing the architectural framework for understanding website IA. Since the web is not a culturally neutral medium, we sought to delineate the cultural dimensions within our formed framework of website IA with the incorporation of the cultural dimensions of Hofstede and Hofstede's (2005), Hall's (1966), Hall and Hall's (1990) and Trompenaar's (1997). This attempt contributes towards the progress of putting a sense of cultural localization to the IA augmentation for local and international website design. In addition, to avoid theoretical aloofness and arbitrariness, practical design presumptions are also reflected.

Keywords: Information Architecture, Culture Interface, Cross Culture, Interface Design, Localization.

1 Introduction

In this paper, we present a theoretical framework on how culture influences to website Information Architecture (IA). Our intention is to bring forward systematic attention by using cultural dimensions as part of theoretically driven approach to the existing framework of website IA. The extension framework of website IA provided in this paper may cater for localization for information products or artefacts where culture dimensions is illuminated alongside the initial architecture framework of website IA. The theoretical building approach is used as a general research method for this study towards incorporating the cultural dimensions into the theoretical framework of website IA, in which inductive reasoning is chosen as part of the theoretical building process [10].

There are few reasons that motivate us to highlight the cultural theoretical framework of website IA. First, IA is one of the areas that has been mostly neglected and still in great need of cross-cultural investigation of web design [4]. The problem also

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 212–221, 2007.
© Springer-Verlag Berlin Heidelberg 2007

lies when IA is highly disregard in cultural specific website [21]. In addition, the development process for website IA may be strongly supported if theoretical grounding support is used as the arbitrary platform in supporting the selection of design methods and principles [8]. However, to avoid the potential danger of arbitrariness and theoretical aloofness, our cultural theoretical framework will include necessary mapping from the cultural framework into practical design indication of IA.

The main objective of this research is to provide cultural theoretical framework into website IA alongside architectural perspectives. It is also our intention that the theoretical framework will contribute towards maximizing user browsing task strategy for information. According to Zhang, Von Dran, Small and Barcellos (1999), browsing tasks are more dependent on web interface designs and thus more congruous to many website designers as comparable to analytical tasks [23]. Zhang et al. (1999) also noted the work of Marchionini (1995) where users' searching behaviours are made up from browsing and analytical strategy, in which analytical strategy is more dependent on the functions of search engines [23]. Furthermore, Kralish and Brenedt (2004) have provide empirical evidence that culture do influence users' search behaviour on website [14]. Therefore, as our research centers on cultural web interface issues, the outcome of our research is leading towards maximizing user browsing task strategy rather than focusing on search engine algorithm issues.

The breakdown of the paper is seen as the following. Section 2 will briefly review our initial framework of understanding website IA using architectural perspectives. Then, basically we incorporate into our initial framework with the cultural theories of Hofstede and Hofstede's (2005) [9], Hall's (1966) [6], Hall and Hall's (1990) [7] and Trompenaar's (1997) [20]. Section 3 illustrates how this framework is used for related IA design practical indication, derived from the cultural theoretical framework of website IA. Lastly, section 4 draws the conclusions and discusses future work and implications.

2 Culture Ascription on Website Information Architecture: A Theoretical Base

For the purposes of our study, we reviewed our initial theoretical framework of understanding website IA derived from architectural perspectives and existing IA literatures [22]. For each of the dimensions, there are theoretical prepositions attached [22]. To incorporate the cultural dimensions into this theoretical framework, we bring forward, the theoretical proposition for 'Context' dimension which is:

"The sense of delightfulness may be imposed, by including the IA elements based on the context or the recognizable sense, in which the underlying surface is used to support the appropriateness of the creation of website IA".

Therefore, the 'sense of delightfulness' emphasized by this theoretical proposition may be imposed with the localization process to the dimension of 'Context', 'Navigation' and 'Content' of IA. This is done by using Hofstede and Hofstede's (2005) [9], Hall's (1966) [6], Hall and Hall's (1990) [7] and Trompenaar's (1997) [20] culture theories as in Figure 1. It is important to use these similar cultural dimensions as the

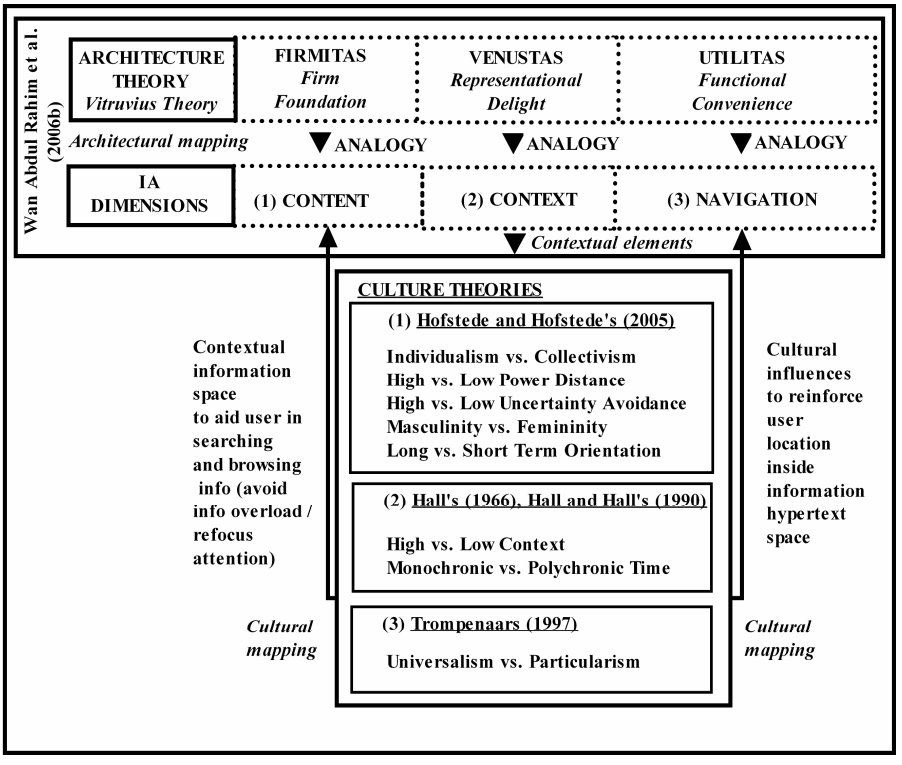


Fig. 1. Cultural Framework of Understanding Website IA (Adapted from Wan Abdul Rahim, et al., 2006b)

benchmark to differentiate and find similarities that exist between cultures. Understanding to differentiate culture and how each culture affects one's behaviors, indicates that we are operating under a different set of expectations [17].

We applied the same method as Khaslavsky (1998) in which to combine a modified framework of cultural values based on Hofstede's, Hall's and Trompenaar's models [12] to determine the degree of cultural adaptation may have influence the dimensions of website IA. The combination approach of using these culture theorists dimensions are used to counterbalance some of the shortcomings noted for each of the models. For example, Hofstede's cultural dimensions have been criticized by some authors as too straightforward and stereotype as the dimension refer not to societies but to nations and employed for different context [14]. Furthermore, the substance of Hofstede's research dates back in the mid '70s and may have changed due to globalization propensity induced by the Internet [2]. In addition, cultural dimensions by Hall and Trompenaars have yet to be empirically validated [18]. Jagne and Smith-Atakan (2006) had made notable criticism to scholars who use these types of cultural dimensions and argued the needs for more studies on indigenous culture [11]. However Hofstede model is extensively replicated and shows a higher level of validity as compared with other alternative models. Therefore, it is being suggested to combine all of

these culture theorists as their works are the most prevalent cultural differentiators and frequently referred to by scholars [3].

There is empirical evidence that culture, at the very least, partially influence web page [3]. Therefore, Hofstede and Hofstede's (2005) [9], Hall's (1966) [6], Hall and Hall's (1990) [7] and Trompenaar's (1997) [20] models were chosen as part of the theoretical building process of website IA. The dimensions were selected by using inductive reasoning methods, which based on related IA practical website design indications to culture dimensions as in Section 3. The explanations of the selected dimensions are as follows:

Individualism vs. Collectivism: The dimensions relate to the relative importance given to individuals or groups within a society. Individualism refers to culture that have loose ties between individuals and on the other hand collectivism in which people integrated into groups and much depended as one cohesive group [9].

High vs. Low Power Distance (PD): The dimensions relate to the state of which weaker member of the society accept inequality in power distribution. Small PD suggests equality; large PD suggests inequality in power distribution [9].

High vs. Low Uncertainty Avoidance (UA): The dimensions refer to the state of a society feels vulnerable of taking risks in unknown situation. Low UA takes risks, whereas high UA are uncomfortable with uncertainty and avoid taking risks [9].

Masculinity vs. Femininity: Masculinity refers to society where gender roles are clearly distinct and femininity stands for society where gender roles overlap [9].

Long vs. Short Term Orientation: Long term orientation encourages virtues oriented towards future rewards whereas short term orientation promotes virtues related to immediate rewards as consideration [9].

High Context vs. Low Context: High context dimension do not require a detailed exchange of information whereas low context require more detailed exchange of information as part of the communication [6][7].

Monochronic vs. Polychronic Time: Monochronic time cultures emphasize on doing things at a time and adhere to rules whereas polychronic time cultures are prone to multitasking and able to adapt changes to initial plan [6][7].

Universalism vs. Pluralism: Universalism cultures emphasize on rules and procedures; pluralism cultures are prone towards relationship based [20].

By identifying culture background based on the cultural dimensions of Hofstede and Hofstede's (2005) [9], Hall's (1966) [6], Hall and Hall's (1990) [7] and Trompenaar's (1997) [20], the study may further uses these cultural theories as guidance to identify design features as part of the localization process of website IA. This may be done with the contextualization process of 'Content', 'Navigation' and 'Context' dimensions of IA which aims on assisting user on searching and browsing for information around the website. In addition, we also propose the contextualization process of 'Content' dimension of website IA that is not only aid user for information searching but also avoid user from experiencing information overload, that occurs when user deals with too much information. As part of the main contribution of this research which is to maximize user browsing task strategy, the contextualization process of 'Navigation' dimension of website IA may reinforce user positioning and orientation while searching and browsing for information [22]. Furthermore, the localization process of 'Context' dimension of IA is also being imposed. Practical design implication can be seen in the next section to see the overall potential cultural effects towards the dimensions of 'Navigation', 'Content' and 'Context' of IA.

3 Practical Implication of Cultural Website Information Architecture

Artefacts are treated as visible and audible patterns of culture which exist on a surface level [1]. Values, on the intermediary level, concern what 'ought' to be done [1]. The dimensions of website IA proposed by our work interplay with these different levels as the values prescribed by the culture dimensions will influence the creation of IA artefacts. The related IA designs that reflect cultural dimensions of influence to the localization process of website IA are accumulated alongside the theoretical framework of website IA. They are based on heuristics and guidelines relevance to website IA design. The attempt was conducted as there are a number of culture interface studies that develop a set of broad cross cultural guidelines as the results which is similar to ones develop by Marcus [13] [15]. The following subsection will highlight few of the existing works initiated on developing cultural guidelines based on Hofstede and Hofstede's (2005) [9], Hall's (1966) [6], Hall and Hall's (1990) [7] and Trompenaar's (1997) [20] culture theories which are related to our dimensions of website IA. We used inductive reasoning methods to ascribe the dimensions of 'Content', 'Navigation' and 'Context' as part of the localization process towards imposing the 'representational delight' of website IA.

3.1 Impact of Culture on Navigation Dimension

Lack of contextual clue is one big problem on web navigation [19]. Hence, culture is treated as a relative attempt to provide situational cues and is manageable by the positioning and selection of elements that is user oriented. The contextualization process for the navigation dimension is oriented towards reinforcing user location inside the information hypertext space. This can be achieved with the contextualization process of sign, icon, symbol, layout, architectural nature (entrance and transition zone) that can further be illuminated with respective cultural dimension of cultural websites [22].

Few IA design relevance to 'Navigation' dimension of website IA is reflected in Table 1. The design prescription is not much concerned about giving advice on how culture can be controlled, but as an attempt of practical relevance of what may be attained by providing constructive ideas for the development of website IA. The design prescription shown in Table 1 is important to create awareness that culture may be an important factor to avoid user from becoming disoriented of their location inside information environment by emphasizing on providing navigation aids. In addition, mapping culture terrain provides guide of how to orient oneself and reduce making errors [1]. Ultimately, it is the patterns, landmark references and cultural nuances that shape the navigation elements of website IA.

3.2 Impact of Culture on Content Dimension

The contextualization process on the dimension of 'Content' of website IA may be achieved with the adaptation of culture theorist cross cultural variables that may have a direct impact on type of labeling, grouping system, colour and typography chosen for the cultural website towards creating the information structure of the website [22].

Table 1. Cultural Indication on Navigation Dimension of Website IA

Culture Dimension	Design Indication
Individualism [9]	- Global and customizable navigational system [16]
Collectivism [9]	- Contextual navigational system [16]
High Uncertainty Avoidance [9]	- Navigation schemes to prevent users from lost [15] - Simplicity with clear metaphors, limited choices and restricted data [15] - Local and contextual navigational system [16] - Include customer service, navigation local stores, local terminology, free trials and download [18].
Low Uncertainty Avoidance [9]	- Less control of navigation; for example, links might open new windows leading away from the original location [15] - Complexity with minimal content and choices [15] - Focus on providing global and local navigation system [16]
Masculinity [9]	- Navigation oriented to exploration and control [15]
High Context [6][7]	- Local and contextual navigational system [16]
Low Context [6][7]	- Global and local navigational system [16] - Links in navigation bar arranged in alphabetical order [19] - Logical and structured layout [19]
Monochronic [6][7]	- Global and local navigational system [16]
Polychronic [6][7]	- Local and contextual navigational system [16]
Universalism [20]	- Global and local navigation system [16]
Particularism [20]	- Local and contextual navigational system [16]

The design inflicted with culture dimensions may deter user from experiencing information overload and may assist in refocusing user attention back to the information structure [22]. Few design relevance on 'Content' dimensions of IA are being highlighted by few researchers as shown in Table 2.

Table 2. Cultural Indication on Content Dimension of Website IA

Culture dimension	Design Implication
Individualism [9]	- Chunk information by task [16]
Collectivism [9]	- Chunk information by modular [16] - Include family theme, clubs or chatrooms, loyalty programs, community relations, symbols of group identity, newsletter and links to local websites [18]

Table 2. (*Continued*)

High Power Distance [9]	- Include hierarchy information and pictures of important people with title [18] - Include quality assurance, awards and vision statements and appeal in pride of ownership [18] - Tall hierarchy in mental models [15] - Highly structured access to information [15]
Low Power Distance [9]	- Shallow hierarchy in mental models [15] - Low structured access to information [15]
High Uncertainty Avoidance [9]	- Mental models and help systems that focus on reducing "user error" [15] - Redundant cues (color, typography and sound, etc) to reduce ambiguity [15] - Chunk information by topic or modular [16] - Include tradition themes, local stores and local terminology, customer service and navigation, free trials and downloads [18]
Low Uncertainty Avoidance [9]	- Mental models and helps systems that focus on understanding concepts rather than narrow tasks [15] - Coding of color, typography and sound maximize information [15] - Chunk information by task [16]
Masculinity [9]	- Include product effectiveness [18] - Clear and distinct gender role [18] - Employ quizzes, games and realism themes [18]
Long Term Orientation [9]	- Content focused on both practice and practical value - Relationship as a source of information and credibility [15]
Short Term Orientation [9]	- Content focused on truth and certainty of beliefs [15] - Rules as a source of information and credibility [15]
Monochronic [6][7]	- Chunk information by task or topic [16]
Polychronic [6][7]	- Chunk information by topic or modular [16]
High Context [6][7]	- Chunk information by topic or modular [16] - Use politeness and soft sell approach in message deliverance [18]
Low Context [6][7]	- Chunk information by task or topic [16] - Use hard sell approach and explicit comparison in message deliverance [18] - Include terms and conditions, rank and prestige and use of superlatives [18]
Universalism [20]	- Chunk information by task or topic [16]
Particularism [20]	- Chunk information by topic or modular [16]

3.3 Impact of Culture on Context Dimension

The cultural framework may add value for localization process of information artefact or information product and can also be used as a design prescription to the construction of 'Context' dimension of website IA [22]. The localization process of website IA will be depended on the culture background of the user and being justified based on the cultural dimensions highlighted in this study. There are few existing design highlighted by related researchers which are relevance to 'Context' dimension of website IA as shown in Table 3.

Table 3. Cultural Indication on Context Dimension of Website IA

Culture dimension	Design Implication
High Power Distance [9]	- Significant and frequent emphasis on the social and moral order (e.g. portrayal of nationalism or religion) and its symbols [15]
Low Power Distance [9]	- Trifling and infrequent use of the social and moral order (e.g. portrayal of nationalism or religion) and its symbols [15]
High Context [6][7]	- Strong preference for visual [19] - Use implicit cultural marker like visual and color [19] - Emphasize on aesthetics value [18]
Low Context [6][7]	- Use explicit culture marker such as page layout [19]

4 Conclusion and Future Works

This research sought to understand and establish the relationship between cultural dimensions to existing architectural framework of understanding website IA. Our goals are to understand what do cultural theoretical perspectives may have on website IA, in which, the understanding and the theoretical proposition highlighted could bring forward valuable knowledge from a known knowledge into IA domain. The study attempt is progressing towards putting localization sense to IA augmentation and implementation. The attempt is justified with theories borrowed from well-known cultural theorist; Hofstede and Hofstede's (2005) [9], Hall's (1966) [6], Hall and Hall's (1990) [7] and Trompenaar's (1997) [20] which are being used to frame up the holistic contextual understanding of the localization process of website IA. An integrated framework, combining these perspectives is presented in Figure 1 as part of the theoretical building process of the website IA framework. This framework was formed by using inductive reasoning research method which is performed by conducting literature analysis on related website IA design to cultural dimensions.

There are several important implications of our research for research and practice. First, we used some existing concepts of website IA to understand this contemporary phenomenon of online information environment. Second, we integrated internal and external perspectives related to cultural paradigm based on cultural dimensions that offer strategically holistic view for website IA design development. This is done by using theoretical propositions suggested by cultural theorist variables to understand

different cultures background by using similar dimensions of comprehensions. Then, this understanding is used as part of the theoretical framework of website IA. In addition, among the implications and contribution of this research is the identification of cultural web design for website IA that may maximize the browsing task strategy for information. The effort may contribute towards increasing the usability level of the website. Furthermore, the practical design indications may avoid theoretical aloofness and theoretical arbitrariness of the theoretical framework of website IA.

There is limitation and constraint imposed by the cultural theoretical framework of this study. The risk is involved as the localization process may fall trap of stereotyping other cultures due to the propensity given to cultural dimension [19] in the development of the cultural theoretical framework of website IA. However, future extension framework may be further evolved by using case study research as part of a theoretical building process [5]. This approach may be useful to support much required research areas that are oriented on engaging culture directly and focused on better understanding of indigenous people [11][19]. In addition, future works may also involve theoretical testing and verification for the theoretical framework. This may be done by using deductive reasoning approach as part of the theoretical testing method and seeks to see if the theory applies to specific instances [10]. Moreover, the framework may be also be operationalized and empirically verified by researchers interested in this area of research. The framework highlighted in this paper may cater for localization process of the information artefacts where culture dimensions may be further illuminated alongside the architectural framework of understanding website architecture. Furthermore, our work could be used as a preliminary point for conducting empirical studies to uncover the dynamics and diverse aspects of IA.

References

1. Alvesson, M.: Understanding Organizational Culture. Sage Publications Limited, Thousand Oaks (2002)
2. Angeli, A.D., Kyriakoullis, L.: Globalisation vs. Localisation in E-Commerce. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 250–253 (2006)
3. Burgmann, I., Kitchen, P.J., William, R.: Does Culture Matter on the Web?. vol. 24. Emerald Group Publishing, pp. 62–76, 1 (2006)
4. Choong, Y.-T., Plocher, T., Rau, P.-L.P.: Cross-Cultural Web Design. In: Proctor, R.W. (ed.) Handbook of Human Factors in Web Design, pp. 284–300. Lawrence Erlbaum Associates Incorporated, Mahwah, NJ (2004)
5. Dooley, L.M.: Case Study Research and Theory Building. Advances in Developing Human Resources (Sage Publications) 4, 335–354 (2002)
6. Hall, E.T.: The Hidden Dimension. Doubleday & Company Inc. (1966)
7. Hall, E., Hall, M.: Understanding Culture Differences. Maine, Intercultural Press (1990)
8. Haverty, M.: Information Architecture Without Internal Theory: An Inductive Design Process. Journal of the American Society for Information Science and Technology 53, 839 (2002)
9. Hofstede, G., Hofstede, G.J.: Cultures and Organization. McGraw-Hill, New York (2005)
10. Hyde, K.F.: Recognising Deductive Processes in Qualitative Research. Emerald Library (2000)

11. Jagne, J., Smith-Atakan, A.S.G.: Cross Cultural Interface Design Strategy. Journal of Universal Access in the Information Society (Springer-Verlag) 5, 3 (2006)
12. Khaslavsky, J.: Integrating Culture Into Interface Design. CHI '98 (1998)
13. Kondratova, I., Goldfarb, I., Gervais, R., Fournier, L.: Culturally Appropriate Web Interface Design: A Web Crawler Study. 8th IASTED International Conference on Computer and Advanced Technology in Education (CATE 2005), pp. 359–364 (2005)
14. Kralisch, A., Berendt, B.: Cultural Determinants of Search Behaviour on Websites. In: Proceedings of IWIPS 2004 Conference on Culture, Trust and Design Innovation (2004)
15. Marcus, A.: Global and Intercultural User-Interface Design. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook, Lawrence Erlbaum Associates, Mahwah, NJ (2003)
16. Mccool, M.: Information Architecture: Intercultural Human Factors, vol. 53, Technical Communication (2006)
17. Mclean, G.N.: Organization Development: Principles, Processes, Performance. Berrett-Koehler Publisher (2006)
18. Singh, N., Zhao, H., Hu, X.: Cultural Adaptation on the Web: A Study of American Companies' Domestic and Chinese Websites, vol. 11, International Journal of Global Information Management (2003)
19. Sun, H.: Building a Culturally-Competent Corporate Web Site: An Exploratory Study of Cultural Markers in Multilingual Web Design. ACM (2001)
20. Trompenaars, F.: Riding the Waves of Culture - Understanding Cultural Diversity in Business. Nicholas Brealey Publishing, London (1997)
21. Wan Abdul Rahim, W.M.I., Nor Laila, M.N., Shafie, M.: Towards Conceptualization of Islamic User Interface for Islamic Website: An Initial Investigation. In: Proceedings of International Conference on Information & Communication Technology for the Muslim World, Malaysia (2006a) http://www.tmsk.uitm.edu.my/ wrahim2/Wan_ICT4M.pdf
22. Wan Abdul Rahim, W.M.I, Nor Laila, M.N., Shafie, M.: Towards a Theoretical Framework for Understanding Website Information Architecture. In: Proceedings of the 8th International Arab Conference on Information Technology (ACIT' 2006), Jordan (2006b) http://www.tmsk.uitm.edu.my/ wrahim2/Wan_ACIT06.pdf
23. Zhang, P., Von Dran, G.M., Small, R.V., Barcellos, S.: Websites that Satisfy Users: A Theoretical Framework for Web User Interface Design and Evaluation. In: Proceedings of the Hawaii International Conference on Systems Science (HICSS 32), Hawai (1999)

Cross-Use: Cross-Cultural Usability User Evaluation-In-Context

Jasem M. Alostath and Abdulwahed Moh Khalfan

Faculty of Computing & Information Systems Department, Collage of Business Studies, PAAET, Kuwait
P.O.Box 23167, Safat, 13092, Kuwait
m_jasem@yahoo.com, dr.khalfan@yahoo.com

Abstract. This paper introduces the Cross-Use experiment, which aims to evaluate the mapping between website design elements and cultural attributes using a user-in-context evaluation approach. This is done by developing three UI designs, and applying them to 63 local participants from the case study cultures (UK, Egypt, and Kuwait). The experiment was conducted using the developed prototypes was able to classify cultures differently, and highlighted those design markers that affects cultural differences in the design of e-banking websites. This is based on user preferences and usability.

Keywords: Culture, Usability, User preferences, e-banking, user-in-context evaluation.

1 Introduction

Many cross-cultural design evaluations use existing websites designs in identifying cultural design differences. However, these design evaluations are not supported with a cultural model, or adopts cultural models that are not design oriented in interpreting design based on culture [5, 6, 7, 8]. In our research of Culture-Centred Design (CCD) we have conducted design evaluations based on the identified subjective cultural attributes (CA) that characterize similarities and differences within and between user groups of different nationality of the cultural model that were developed based on HCI design [9, 13]. The most important advantage of this new approach is that the results of the analysis provide the designer with sufficient information to generate new websites that are more sensitive to culture and genre variability. However, the designs generated are not guaranteed to be optimal. This is because: (1) the existing websites that form the basis of the analysis may not have been well designed from the cultural point of view, (2) the claims from the cultural-design mapping from which designs are generated may be insufficient to determine a unique design decision, and (3) the design analysis that is undertaken does not provide any important information on design aspects such as usability [9]. Our solution to this problem is in the CCD methodology [9], which uses the design analysis results to develop a number of possible prototype websites that will be culturally adapted to some degree. Then a rigorous user testing approach is used to decide between the alternatives (further details about the CCD method see Alostath [9]).

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 225–234, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 Cross-Use: Method and Process

The experiment design involves three national cultures, using three user interfaces for simple and complex tasks (3*3*2 mixed design). The independent variables of the cultural factors were manipulated using three designs and are shown using the Latin Square design to counterbalance order effects [1]. The prototype used in this experiment was developed from scratch by the researchers based on the results of the design analysis. The three websites developed have one user interface design for each culture that maximizes the cultural and genre attributes appropriate for that culture. In addition, for each of the interfaces developed design alternative with content that is appropriate for each of the other cultures being tested is also included. This is done by exploiting the XML technology[1].

2.1 Variables and Participants

84 user variables are measured in this experiment. Fourteen variables are required to collect participants' demographic information. Of the remaining 70 variables, 58 are the users' subjective valuations of interface properties (e.g. text, images, and others) that are thought to have a cultural impact. The remaining 12 variables are used for evaluating each group of tasks (simple and complex tasks). Each group has six variables, of which four measure usability and two measure culture and trust compliance. These six variables are repeated for each task group. These 12 questions are aimed at building a usability factor that can be used to determine: (1) at the high level, the most usable design for each of the studied cultures, and (2) at the lower level, the design markers[2] (DMs) that improve usability from the 58 DMs. The experiments were conducted with 21 participants from each culture (Kuwait, UK, and Egypt). Participants were selected based on their ability to use the computer, internet, speak English and were given financial incentive.

2.2 Procedure and Materials

The Cross-Use experiment procedure consists of seven stages as shown in Figure 1. In the first stage, participants were informed about the three experimental sessions, objectives and procedure, and were required to sign the consent form. This is followed by the second stage, where each participant receives two 3-digit personal account codes and a password that allows them to run the experiment process and perform the online transactions required.

In the third stage, a questionnaire of 28 questions is administered; each question included one or more images of a DM relevant to one of the design claims being investigated. The aim is to obtain an initial understanding of the participants' expectations before interacting with the e-banking prototype. In the fourth stage (Task performance evaluation), the participant starts to perform six tasks, which are divided

[1] XML usualy used to display different data across different UI platforms (e.g. Computer UI, mobile interface and others). Here, it is used to display different cultural data into HTML file, and this is based on users' culture.

[2] Design marker is a concrete design aspect and its existence is expected to have a cultural or genre, or other relations.

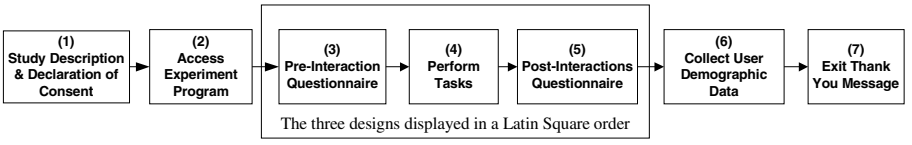


Fig. 1. Cross-Use experiment procedure

into two task groups (simple and complex tasks). Each group contains three tasks, the first three are for information inquiry and the other three are for performing transaction tasks. Upon completion of the three tasks, a comparison questionnaire is administered to rank the tasks. After each of the three tasks, participants answer the six design comparison questions, which compare the three designs in terms of usefulness, ease of use, frustration, satisfaction, culturally related issues and the most trustable design. The aim of this stage is to obtain the most usable design and what are the DMs that make a design usable for a particular culture. In the fifth stage, the participants were presented with several design layouts, and transactions processes necessary to explain the question, and were asked design-specific questions to rank several cultural design claims (30 questions presented in a forced-choice comparisons as well as 5-point Likert scale questions). The aim of this stage is to measure users' experience after their interaction with different interface designs and performing different types of tasks. The final stages are used to wrap-up the experiment by collecting participants demographic data and ending with a thank you message.

The experiment uses a Pentium Centrino 1.5 MHz laptop with 15" TFT screen, and regular mouse. The experiment was executed from the local web-server running on the same computer. In addition, a reasonable resolution (320 x 240 pixels) webcam was connected to the computer to record the participants' facial expressions using Morae™ tasks recording tool (see www.techsmith.com).

2.3 Objectives and Hypotheses

The objective of the Cross-Use experiment is to substantiate the cultural design claims [9, 12], which have been substantiated earlier in design evaluations approaches [9]. This experiment further substantiates these claims based on user-in-context evaluation, and aims to provide two types of results. These are related to the user preferences, and usability for the selected design, and design markers. User preferences refer to the results based on a comparison made by the user between two or more UIs or on specific aspects of those designs. In contrast, usability is assessed by performing real tasks, and then both objective (e.g. time to perform a task) and subjective (e.g. satisfaction with task) outcomes are measured. The results of users' preferences and usability are also useful in deciding whether the design preferences are a good indicator for usability. In order to test these objectives, several analysis methods were conducted, to examine the validity of the following hypotheses:

H1: When given a choice between a website designed for a different target culture and one designed for their own target culture, users will prefer the website designed for their own culture.

H2: Websites that have been designed for a particular target culture (e.g. Kuwait, or Egypt, or UK) using the developed cultural design claims will produce better usability results when tested by members of that particular target culture.

H3: Using Discriminant Analysis (DA), it is possible to identify specific or aggregated DMs that are the main contributors to the observed user preferences and usability improvement.

In this study, the DA and Chi-Square statistical analysis methods were used to analyse the questionnaire data, which involves a 189 observations -- 63 observations for 3 designs. The DA is used to show the most important or interpretive independent variables, which discriminate the dependent variable or affect it [11], while the Chi-square is used to determine whether the groupings of cases on one variable are related to the groupings of cases on another variable [2].

3 The Cross-Use Experiment

The aim of the Cross-Use experiment is to present the important DMs that were identified by users' preferences, and usability. This can be determined by two analyses, which are concerned with the ability of the developed user interface designs to classify the cultures differently, and the identification of those DMs that play a significant role in causing these differences. The key factors in this analysis are usability and preferences.

3.1 Cross-Cultural Design Preferences

Study hypothesis (H1) predicted that when creating designs that are in accordance with cultural design claims [9], these designs are able to generate culturally sensitive designs. The data collected from the experiment were used in this analysis to classify the three cultural groups of users according to their preferences for the identified cultural designs. DA was performed with national culture as the dependent variable, and the DMs as independent variables. The results of this analysis confirmed hypothesis H1 (see Figure 2 and Table 1). This indicates the ability of the website designs that adopted the cultural design claims to design for different cultures to capture users' different preferences. The DMs that cause the cultural preference differences among specific national cultures resulting from the above DA test are shown in Table 1.

3.2 Cross-Cultural Design Usability

In this section, an investigation of a good representative score for the cultural usability factor is conducted. Then, two types of analysis are performed. The first analysis uses a Chi-square test, and the second uses DA. The first analysis tells whether or not there is a relation between national cultures and design usability. The second analysis helps in classifying designs according to cultural usability and DMs, and identifying the DMs that are used to improve usability for each culture.

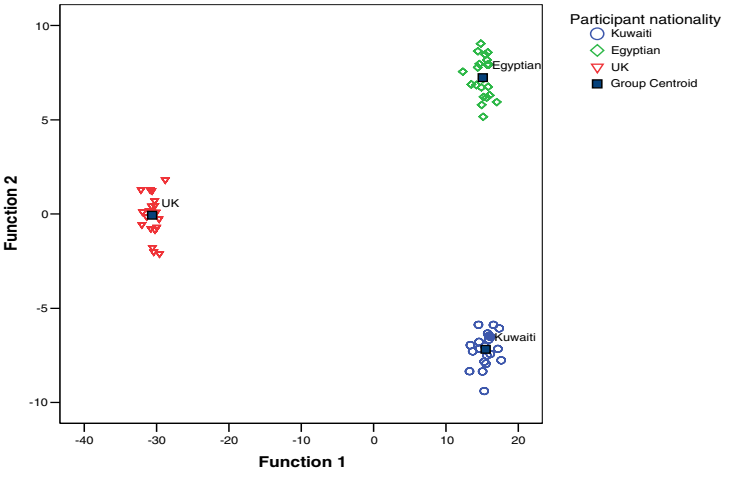


Fig. 2. Canonical Discriminant Functions plot: visualizing how the two functions discriminate between cultural groups by plotting the individual scores for the two functions

Table 1. Partial summary table for the user preferences DMs

CA	Claim	Design markers	KU	EG	UK	Related Question
		Relationship Metaphors				
R6, R7	C16[3]	Religious Metaphors (Design A)	M	M	L	B2a (*)
		National Metaphors (Design B)	M	H	M	B2b (*)
		Neutral Metaphors design (Design C)	H	H	H	B2c
		Navigation tools				
T4	C21	Drop-down Menu (complex navigation)	H	M	H	A1a (*)
		Tree-view (complex navigation)	L	M	L	A1b
		Sense of security				

Legend
CA is refer to the cultural attribute code identified in the HCI-cultural model [see 10]
- Low (L): <2.49; Medium (M)=2.50..3.49; High (H): >3.49
- (*) DM identified to be significant (p<.001) based on both the DA with Univariate ANOVA tests
- No sign indicates the DM was significant based on DA (p<.001) but not significant across cultures based on the Univariate ANOVA test (p<.001).

Culture and Usability Relation. The aim of this analysis is to identify the design differences affecting usability among the three cultures, based on the usability factor. Here, attempts are made to find if there are any relationships between national cultures and design usability. If there are any, then the DMs that are affecting usability across

[3] Claim (C16): High racial tendency oriented cultures (relationship) are expected to show high use of religious and/or national symbols in the design more than low racial tendency oriented cultures, which tend to show neutral symbols.

these cultures are investigated. The study hypothesis (H2) predicts that when creating designs for cultures based on the cultural design claims and design investigation results (presented in [9]), such as design (A) for Kuwait, design (B) for Egypt, and design (C) for UK, such designs are expected to show better usability results by members of those particular cultures in their own cultural designs. Based on study hypothesis (H2), two issues need to be verified: the first issue is in determining whether a relation exists between culture and usability, which was verified using a Chi-Square test. Then, the second issue is determining the usability improvement that occurs frequently within the targeted cultural design, which was verified using a DA test.

As for the existence of a relation between the design usability (represented by the usability factor) and the national cultures, the following hypothesis was defined:

> **Hypothesis:** There is a relation between national cultures and designs' usability (dependent)

A Chi-squared analysis shows that there is a significant relation between national culture and design usability (χ^2=19.08, df = 4, Sig. < 0.001). In Figure 3, certain website designs are found to be more usable by certain national cultures is shown.

In validating hypothesis (H2), which predicted that websites that have been designed for a particular target culture (e.g. Kuwait, or Egypt, or UK) using the cultural design claims will produce better usability results when tested by members of that particular target culture. Figure 3 shows a clear tendency for high usability by Kuwaiti participants in using their cultural design (design-A), but there is an exception to the hypothesis for Egypt and UK. Egyptian participants show high usability in using design-A, while UK participants have a usability score that is split between design-B and design-C. To further investigate the cause of this unexpected result, in the following section, the DA is used to identify which specific variables were affecting usability scores for each of the cultures.

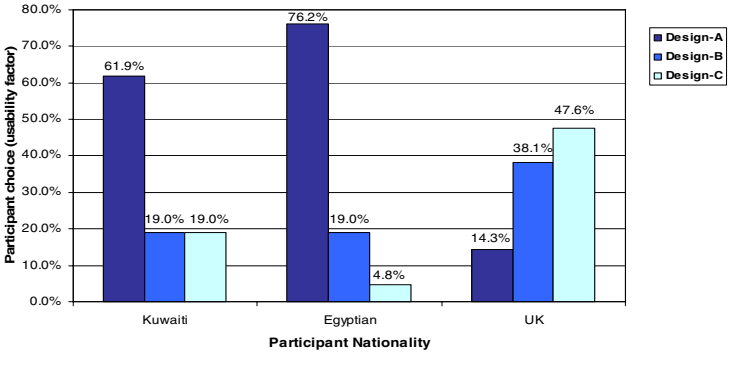


Fig. 3. The distribution graph for the usability scores according to culture and design

The Classification of the Three Designs Using DA Test. DA was performed with usability factor as the dependent variable, and the studied CMs (58 variables) as independent variables. This test provides two types of result. The first result is the classification of the three designs (A, B, and C) based on the usability factor for each

case study culture (to determine the usability level on different designs). The second result is in identifying the DMs, which cause usability improvements among specific national cultures as shown in Table 2. The DA results shows that the total validity of the proposed model is 100% for observations, which indicates that all cases were adequately categorized in all cultures. In addition, the visual graphs produced by the DA [9] show a divergence between the design type centroid points, which primarily discriminate between UK, Kuwaiti and Egyptian cultures. However, the design classification based on usability factor across cultures shows that design-A seems not to discriminate between Kuwaiti and Egyptian cultures. This confirms the results shown in Figure 3, which stresses that at the cultural usability level, Kuwaiti and Egyptian participants show some similarities in usable DMs. This indicates that, based on usability, Kuwait and Egypt could share design-A and that the UK site (design-C) should be redesigned to have cultural DMs from design-B, in addition to design-C DMs. Thus, study hypothesis (H2) is partially confirmed for Kuwaiti culture. However, to be sure of this conclusion we need to look at the DA results in more detail in order to determine which particular design factors were causing these usability effects. This will enable us to determine how to fine-tune the designs and modify the identified cultural design. The specific details of the DMs that affect these changes are identified and discussed in Table 2.

As can be seen from the summary DA results shown in Table 2, there is a clear tendency to identify specific DMs that are the main contributors to the observed participants' usability. Hence, H3 is confirmed for identifying the DMs for usability. This indicates the ability of the DA to identify the DMs that affect usability. These DMs are used as user-in-context based evidence in supporting or contradicting the cultural design claims. Reviewing the complete list of the usability DMs (see [9]) indicates that the shared DMs and cultures based on the cultural usability factor shows that there are more shared cultural usability DMs between Kuwait and Egypt, followed with Kuwait and UK. However, between Egypt and UK, there are no shared DMs. Again this confirms the relation between Kuwaiti and Egyptian cultures discussed earlier in sections 3.1 and 3.2.1. In addition, the DMs related to preferences and usability levels, the analysis shows that the identified DMs for preferences are higher than usability (see [9]). Furthermore, some usability markers appear to be different from preferences related DMs.

Table 2. Partial summary table for cultural usability DMs

CA	Claim	Design marker	KU	EG	UK
		Relationship Metaphors			
R6, R7	C16	National Metaphors (Design B)		H†	
		Navigation tools			
T4	C21	Drop-down Menu (complex navigation)	H†		H†
		Tree-view (complex navigation)			L†
		Drop-down field (complex navigation)	H†		H†
		Free-search (complex navigation)		H†	

Legend

† This symbol indicates that this DM affects usability for this particular culture (presenting a cultural-usability design). The result of this indicator is determined by performing DA.

4 Discussion and Conclusion

The Cross-Use data analysis was presented through two models. The first model is the cultural preferences model, which consists of the high level classification and DMs of cultural preferences (as shown in Section 3.1), and the second model is the cultural usability model, which consists of the high level classification and DMs of cultural usability (as shown in Section 3.2). Both models have different concepts that require various analysis techniques, which produce diverse results and significance levels. The cultural preferences model concept was to identify whether the participants' preferences for using the three designs are different, where the experiment shows there are significant differences. This proves that the experiment designs were able to classify cultures based on participants' preferences for the DMs, which at one level substantiates the experiment design and on the other level shows that there are cultural design differences. In addition, this model shows that a high number of the identified DMs are culturally preferred, which indicates that most of the DMs can be differentiated based on participants' preferences.

The next challenge here was to see whether the usage of culturally preferred DMs in local designs improves local design usability. This led to the development of the second model, which covers usability and was referred to as the cultural usability model. The cultural usability model was developed based on how the user performs the assigned six tasks (see Section 2.1), where the usability factor was developed to discriminate between the studied cultures. Based on this model, several issues were identified. The first issue shows that there is a high relation between culture and design usability using the three designs. This indicates that the three designs were able to identify a relation between culture and usability, which shows that at the classification level culture preferences are able to make usable designs. However, based on the most usable design related to culture, the results show that the Egyptian culture reflects design-A as the most usable design compared to the earlier expectation, which is design-B. In addition, the UK participants shared both design-C and design-B as they are the most usable designs (as shown in Figure 3). Therefore, the cultural DMs based on usability are not the same as the cultural design claims. These findings motivate the investigation of cultural usability DM.

Earlier, design preferences and usability were discussed to determine their differences. Then, during the experiment evaluation, these two issues were tested using a process to evaluate users. The question here is whether the websites that have been designed based on user cultural preferences are necessarily presenting usable design. The answer to this question helps in recognizing the sensitivity of the approach in collecting data that provides results to help in delivering usable design. The study of Evers and Day [3] uses the culturally extended Technology Acceptance Model (TAM), which uses the usability variables such as usefulness, ease of use, and satisfaction to determine the UI acceptance. They use questionnaires to collect users' preferences. Their study indicates that design preferences affect interface acceptance across cultures. In the Cross-Use experiment, the general view of the design classification based on the usability factor for each culture shows higher differences on cultural preferences than usability (see [9]). This proves that participants prefer design differently, but when they use the design, it shows more differences in usability than

originally expected. This highlights the complementary usage of the user-in-context evaluation in determining the usable cultural DMs.

Many website developers and evaluators use methods that assess user preferences aiming to create usable design. For example, the Cultural Markers [5], Website Audit [8], and user evaluation [10] using questionnaire based tools only are not sufficient in understanding and identifying the appropriate usability requirements. According to the results of Cross-Use experiment, as can be seen from Table 1, which presents user preferences CMs, and Table 2, which presents usability CMs, the comparison between the two markers indicates that the number of the identified markers in each type is different, and the identified markers based on preferences are not necessarily identified based on usability and vice-versa. The cultural usability model identifies fewer DMs than in the cultural preferences model. These prove that not all of the preferred DMs are necessarily usable DMs. Furthermore, the cultural usability DMs show that there are some DMs that are not shown to be preferred by the participants but are statistically proven to improve usability (e.g. Tree-view navigation DM in claim C21, as shown in Table 1 and Table 2). This suggests that research based on design preferences does not necessarily present the effects of usability as indicated by Constantine and Lockwood [4]. As a consequence, the results of such studies linking participants' preferences to design can be doubted, and this also affects the investigation of existing website design, as both adopt the same results. Therefore, the results obtained from users' preferences and usability should scale differently in supporting cultural design claims and in the later stages of the development of cultural design guidelines.

This conclusion strengthens the research results as they are obtained by evaluating both the cultural preferences and usability DMs. For the future research a detailed inspection method are expected to be used to analyse these results together with results of earlier research studies, which aims at developing evidence-based cultural design guidelines and recommendations.

References

1. MacKenzie, S.: Research Note: Within-subjects vs. Between-subjects Designs: Which to Use? Toronto, Ontario, Canada, http://www.yorku.ca/mack/RN-Counterbalancing.html (2002)
2. Pallant, J.: SPSS Survival manual. McGraw, New York, USA (2005)
3. Evers, V., Day, D.: The Role of Culture in Interface Acceptance. Human-Computer Interaction, Interact'97, London (1997)
4. Constantine, L., Lockwood, L.: Software for Use: A Practical Guide to the Models and Methods of Usage-Centred Design. Addison-Wesley, New York (1999)
5. Barber, W., Badre, A.: Culturability: The Merging of culture and usability. In: Proceedings of the 4th Conference of Human Factors and the Web (1998)
6. Marcus, A.: User Interface Design and Culture. In: Aykin, N. (ed.) Usability and Internationalization of Information Technology, pp. 51–78. Lawrence Erlbaum Associates, Inc, Mahwah, New Jersey (2005)
7. Bourges-Waldegg, P., Scrivener, S.A.R.: Applying and testing an approach to design for culturally diverse user groups. Interacting with computers 13, 111–126 (2000)

8. Smith, A., Dunckley, L., et al.: A process model for developing usable cross-cultural websites. Interacting with computers 16, 63–91 (2004)
9. Alostath, J.: Culture-Centred Design: Integrating Culture into Human-Computer Interaction. Doctoral Thesis, The University of York, UK (2006)
10. Evers, V.: Cultural Aspects of User Interface Understanding: An Empirical Evaluation of an E-Learning website by International User Groups. Doctoral Thesis, the Open University (2001)
11. Brace, N., Kemp, R., Rosemary, S.: SPSS for psychologists: A guide to data analysis using SPSS for Windows, vol. vii, p. 287. L. Erlbaum Associates, Mahwah, NJ (2003)
12. Rosson, M.B., Carroll, J.M.: Usability Engineering: Scenario-Based Developement of Human-Computer Interaction. USA (2002)
13. Alostath, J., Wright, P.: Integrating Cultural Models into Human-Interaction Design. Conference on Information Technology in Asia 2005 (CITA2005), Kuching, Sarawak, Malaysia (2005a)

Testing Remote Users: An Innovative Technology

Rebecca Matson Sukach Baker[1], Esin Kiris[1], and Omar Vasnaik[2]

[1] CA One CA Plaza
Islandia, NY 11749, USA
[2] Microsoft Corporation, One Microsoft Way
Redmond, WA 95052 USA
rebecca.baker@ca.com, esin.kiris@ca.com,
omar.vasnaik@microsoft.com

Abstract. Conducting usability tests with remote users require unique approaches and techniques. Remote users often have requirements that differ significantly from local users, as the technology is not wholly contained in a controlled usability lab. Based on the authors' experience with remote usability techniques, this paper provides insights and practical tips about a technology used to host and monitor remote usability tests, users' reactions to the testing technology, and the communication rhythm within the testing organization.

Keywords: Remote Usability, Usability Test, User-Centered Design.

1 Introduction

Remote usability tests are becoming more main stream today. Companies are now developing products and solutions for global users in a global workforce. Gone are the days when it would suffice to get feedback on product designs from users "down the hall." Getting early user feedback from an international audience reduces project schedules and, consequently, costs, and perhaps more importantly removes cultural bias in product design.

Several remote usability tools exist in the market today. But experimentation with these tools has shown that a single all encompassing solution to run effective remote usability tests does not exist. There are several factors to be considered to successfully running formative and summative remote usability tests,

- Can the test evaluator watch users interacting with prototypes in real time?
- Is there a security mechanism in place to authenticate users and prevent screen scraping (an issue where screenshots are taken without the test evaluator's knowledge)
- Can the test evaluator record in real time a user's voice, mouse clicks, and keyboard strokes?
- Is the recording in true color (uses 64-bit color or higher) with the audio and video integrated?
- Can highlight videos be easily made and distributed?

Vasnaik and Longoria [1] provided a remote usability infrastructure solution which addresses some of the issues stated above. They combined a number of tools (Citrix

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 235–242, 2007.
© Springer-Verlag Berlin Heidelberg 2007

GoToAssist[1], Windows[2] Active Directory, Windows Live Meeting, and Windows Net Meeting) to produce a comprehensive remote solution. This solution although complex in nature worked well for many but not every situation. The solution was cost effective and provided good data. However, while it worked well for thin-client solutions (prototypes on the web) it could not be used for live thin-client and thick client (Windows, Unix[3], etc.) applications.

Usability testing live thin and thick client applications forms an intrinsic part of the Software Development Lifecycle (SDLC), especially for enterprise organizations. Bringing customers' onsite or traveling to customer sites is a very expensive proposition. Enterprise organizations have customers widely dispersed and work on product offerings that are complex to install and set up.

Brush, Ames, and Davis [3] and Hartson, Castillo and Kelso [4] have determined that significant differences between remote usability tests (both synchronous and asynchronous) and traditional tests do not exist, in terms of the number of usability issues found, their types, or their severities.

The case study presented here builds on the work done by Vasnaik and Longoria [1], providing a technical solution where thin and thick client applications can be remotely tested.

2 Remote Usability: Challenges and Requirements

Testing thick client applications remotely offers challenges that are logistical, political, and technical in nature. We needed a methodology that would meet the following requirements:

- Require minimal installation of software on users' machines
- Is accessible to users within their company firewalls (the solution should be accessible to both the test evaluator and the end user)
- Allow the test administrator to view user actions in real-time
- Record voice and screen interactions (mouse clicks and keyboard strokes) simultaneously and store those recordings as a single integrated file
- Host multiple sessions of the software application being tested on different operating systems
- Be maintained by technical personnel within the authors' organization, outside of CA's User-Centered Design (UCD) group
- Have 24x7 technical support for the test administrator
- Is accessible to developers for installation of test software – that is, it allows installation of software to be done by individual development teams rather than requiring UCD team members to perform software installations.

After testing several applications, we decided to use a combination of Citrix hosting software, Unicenter Remote Control[4], Microsoft Live Meeting, and VMWare[5].

[1] Trademark or registered trademark of Citrix Corp.
[2] Trademark or registered trademark of Microsoft Corp.
[3] Trademark or registered trademark of The Open Group.
[4] Trademark or registered trademark of CA, Inc.
[5] Trademark or registered trademark of VMWare Inc.

Citrix was used to create a secure hosting solution outside the firewall. Unicenter Remote Control was used to allow users access individual VMWare sessions. By using VMWare, we were able to host multiple sessions with different operating systems and applications – Unix, Windows, etc. sessions could all exist on the Citrix server simultaneously. In addition, the VMWare sessions could be created on a separate server that developers could access to install software before the sessions were uploaded to the Citrix server. Unicenter Remote Control allowed us to view the users' interaction with the VMWare sessions in real-time. Microsoft Live Meeting was used to record the voice and screen interactions simultaneously and store the result in a single integrated file that could be downloaded. Keyboard strokes and mouse click data were not collected.

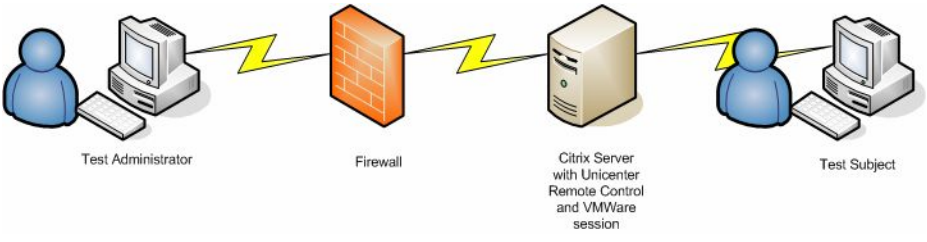


Fig. 1. Remote Usability Test Environment

The Customer Service Center at CA took over responsibility for maintaining the Citrix server. They agreed to provide technical support for the test sessions as they were scheduled, ensuring someone was on call during the test in case any technical issues arose.

3 Case Study: An Enterprise Network and Systems Management Application Remote Usability Test

The product chosen for our case study was an enterprise network and systems management thick-client application. (Windows and Unix). This product is an interface designed for an enterprise administrator that manages enterprise systems and applications from a single user interface. The first step in evaluating this product involved conducting a usability test of the current application to understand the usability issues users faced while interacting with the application. The target users were enterprise administrators located in the Netherlands, Turkey, and the United States.

3.1 Finding Users and Design Partnership Program

The most important challenge of usability testing process is to recruit users from different enterprises. This can be particularly difficult since users are busy with their jobs and often have no time to participate. From this perspective, we are lucky to have very active user pool. We started a program called Design Partnership program (https://connectprimary.ca.com/webac/wac/usabilitylogin.asp). The goal of this program is to

reach our users by soliciting their feedback and experience when using our products. We made users understand that they form an integral part of the design process. The Design Partnership program was one of the many channels we used to reach users for this test. We also worked close with product management and training teams to reach users.

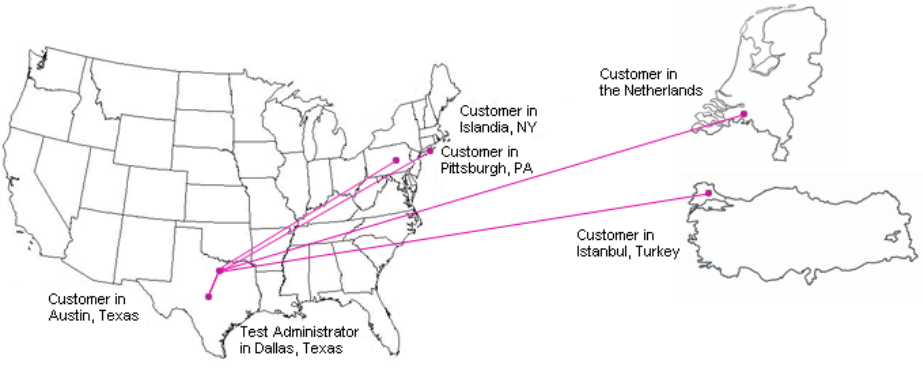


Fig. 2. Geographical distribution of users and test moderator

Eight users from four customer organizations participated in the remote usability test. These users were located in five different locations including the Netherlands, and Istanbul, Turkey. The use cases being tested required users to have a system administrator background, which in turn limited the number of users we had accessible to us. We contacted the users by email. Participation was strictly voluntary and no incentives were provided. In addition, we asked users to sign an Informed Consent Form prior to the test. This form stated that we had permission to record the users and use the data for product improvement. It further guaranteed their anonymity as a tester.

3.2 Test Preparation

To prepare for the tests, we had to:

- Create a schedule of the tests for Customer Service Center to ensure availability of support. The Customer Service Center administered the hardware and hosting software for our tests. As a result, we needed to ensure they were available during the tests in case an issue arose due to the infrastructure.
- Create and configure a VMWare session. A VMWare session is an image – or virtual representation – of a machine. This image could be a Windows NT server or a Linux PC, etc. The Customer Service Center provided generic images for our use of any operating system/hardware combination we required.
- Arrange for the development team to install the enterprise software on the VMWare session. We uploaded the image we got from the Customer Service Center to a server and provided basic communication information (such as IP addresses) and software (such as Unicenter Remote Control). We then contacted the

development team and had them install the enterprise software to ensure that it was configured correctly.

- Provide the configured VMWare session to the Customer Service Center to be uploaded to the Citrix server. Once the session was configured with the software to be tested, we contacted the Customer Service Center and had them upload the session onto the infrastructure outside the firewall.
- Create a session in the Remote Control local address book for both the testers Citrix ID and the users Citrix ID. Remote Control allows users to access computers (in this case, VMWare sessions) on remote machines. You must either know the IP address of the machine you wish to access or you must have shortcuts set up in a local address book. We created shortcuts to make accessing the VMWare session through remote control easier.
- Run a pilot test on the Citrix server to validate test steps, test language, and response time.

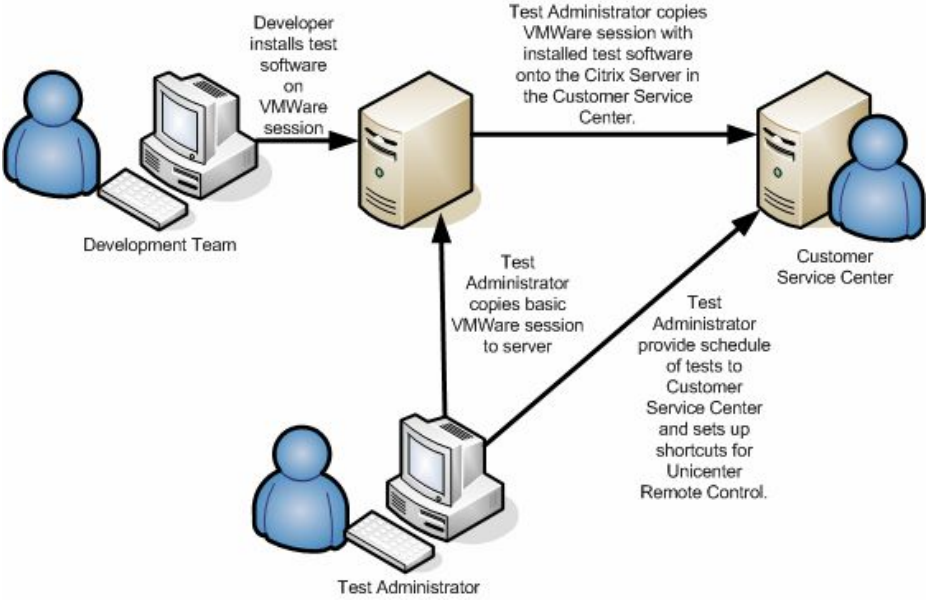


Fig. 3. Remote Usability Setup

3.3 User Preparation

Before the test, users received an instruction email that provided them the time and date of their test. They were asked to have the following items ready before the test started::

- Access to a computer with:
  - Internet Explorer 5.5 browser or higher
  - Java 1.4 or higher

This software was necessary for the user to be able to access our Citrix software.

- Access to a speaker phone or a headset, with toll free access. Because the test required the use of a mouse and keyboard, we asked that the users have a way to communicate with us without encumbering their hands.
- A printed copy of the task list. We wanted the users to access the software in a full screen mode, which made it necessary for them to have a printed copy of the tasks.
- A copy of the Informed Consent Form faxed to the test administrator. Because we were recording the sessions, we needed to have permission from the testers. The Informed Consent Form provided notification that they were to be recorded and gained their permission, in a legal fashion for the use of that recording for product development purposes.

3.4 Test Instructions

Users were asked to follow these steps to connect to the test:

1. Using a speaker phone or headset, dial into the following conference number:
2. At this point, they were greeted by the test administrator who confirmed that we had received the informed consent form and walked the user through the remaining setup steps.
3. Go to http://www.ca.com/wwsolutions and logon using the following information:
   UserID: xxx01
   Password: wwxx
4. Double click the URC icon in the applications box.
5. If prompted for access to local user files, Choose No Access to Client Files and Never Ask me again. Click OK. Remote Control Explorer appears.
6. Expand the Viewer branch in the left hand tree and click Local Address book. On the right, double click the ***SESSION NAME*** icon.
7. A Connect screen appears. Do not change any of the information on this screen. Click Connect.
8. On the toolbar, select the computer icon to switch to full screen mode for the test.

3.5 Test Results

Once users accessed the product VMWare session, the test moderator started the recording. This involved recording the users' screens and the audio conversation. The test then proceeded like a traditional usability test. During the test, users performed seven tasks over the period of an hour. No difference between the performance of the test application and the actual application were found – two users made the comment that the test application ran more quickly than the one they had installed on their local machine. Recordings of the test had good visual resolution and excellent audio quality. The test administrator was able to review the test sessions while analyzing the test results and post the recordings on an internal company website for review by the development team. Support was needed from the Customer Service Center once – the problem was an issue with the user's machine that was quickly resolved.

4 Results: Advantages and Disadvantages

Advantages and disadvantages of remote testing over traditional usability lab testing have been summarized in previous studies[1], [2]. They found that remote testing provided significant savings, reduced the amount of time required to arrange the tests, provided a greater diversity in the user base, and increased the comfort level of the testers. They also found that some users were concerned with security issues, that the technological compatibility of systems could cause problems, that users could become distracted by their environment or be compromised by reading tasks early, that the recording quality was not as high as that found in labs, and that the inability to physically observe the users may cause loss of data.

This technique improved on previous studies by addressing both the issues of user security and technological compatibility. In addition, it provided a method of testing thick client applications with the same alacrity as thin client applications.

This method did not address the issue of "screen scraping" (users taking screenshots of the interface). This was not an issue for this test as all users were existing customers with an ongoing relationship with CA.

5 Cost Analysis

The following traditional usability evaluation cost exercise is adapted to this paper from the previous paper written by Vasnaik and Longoria [1]. A previously completed traditional usability evaluation of another enterprise application was analyzed in terms of cost. One UCD group member visited two domestic customer locations. Only eight users could participate in that evaluation. The entire exercise took about ten business days and travel costs alone were about $5000. Because of prohibitive travel costs only a single UCD group member could travel to the customer sites and the product team could not participate in the process.

Eight users in five customer locations spread across three continents participated in the remote usability evaluation of the thick client applications. This exercise was completed in one and a half business days. If this test had to be conducted traditionally it would have taken about 6 business days. Travel cost for the test moderator and product manager would have been in the region of $15,000 which is a conservative estimate. This remote usability evaluation resulted in shortening the development cycle by two weeks and significant monetary saving. The additional time was better utilized to perform more design iterations and test with additional users. In addition, the remote usability evaluation enabled testing the interface with a diverse user population including international users. In addition, the initial expenditure for the remote usability infrastructure was negligible when compared to maintaining a traditional usability lab.

6 Conclusions: Findings and Future Research

This case study showed the feasibility of the proposed remote usability technology, from technological, logistical, and political standpoints.

Logistically, users found little difference between using the application on their local machines and using it remotely on the test machine. Further studies need to be done to determine how and whether the results from remote tests such as these vary significantly from tests performed in a lab setting [3].

Politically, reliability of support from other teams within CA was very good but could be improved by providing standard instructions for processes (how/when to contact the Customer Service Center, how developers can access the VMWare sessions, etc.). By formalizing the processes, we expect to be able to establish solid expectations between groups, and hence a greater sense of responsiveness and responsibility.

Technologically, the applications provided a stable environment for testing with few integration issues. Due to the relatively large number of applications used to create the environment, the initial learning curve for UCD members to set up and run a test was higher than desired. Standard VMWare templates and detailed instructions helped to some degree, but further training materials would improve this area.

Acknowledgements. Mike Belli from the Customer Service Center, Samier Nadji from the Unicenter NSM development team, and Kerry Harrison from the UCD team at CA, Inc.

References

1. Vasnaik, O., Longoria, R.: Bridging the Gap with a Remote User, UPA (2006)
2. Kauss Methodology for Remote Usability Activities: A Case Study IBM Systems Journal (2003)
3. Brush, A., Ames, M., Davis, J.: A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface. In: Proceedings of CHI ’04 Conference on Human Factors in Computing Systems, pp. 1179–1182 (2004)
4. Hartson, H.R., Castillo, J.C., Kelso, J., Neale, W.C.: Remote Evaluation: The Network as an Extension of the Usability Laboratory. In: Proceedings of CHI ’96 Conference on Human Factors in Computing Systems, pp. 228–235

Web Usability and Evaluation: Issues and Concerns

S. Batra[2] and R.R. Bishu[1]

[1] Department of Industrial Engineering, University of Nebraska-Lincoln, Lincoln, NE
[2] Cognitive Ergonomist, Enterprise Rental Inc., St. Louis, USA

Abstract. This paper presents a summary of usability work done at the Usability Laboratory at University of Nebraska in the last few years. The main objective of the first study was to compare the efficiency and effectiveness between user testing and heuristic analysis in evaluating four different commercial websites. The results showed that both user testing and heuristic analysis addressed very different usability problems and both methods are equally efficient and effective. In the second study. the primary purpose was to compare the performance between remote usability testing and traditional usability testing. The results indicate that remote usability testing is no different from traditional usability testing. The third study attempted to look at cultural differences in web usability. The results indicated that cultural dimensions have significant effects on user's web preferences. The primary objective of final study was to determine if user's surfing behavior could be predicted through their cognitive style. Results show that cognitive span scores are not strong enough to form association rule with individual difference clusters of web surfing behavior. The results are discussed with respect to all perspectives of Web.

1 Introduction

The Last three decades have seen the communication field go through a technological revolution. Internet has changed professional and domestic life completely. Web based applications have become a standard, cross-platform, nonproprietary means for businesses to communicate with each other and with consumers. This in turn has made the design of a website and a web interface very important and an integral part of contemporary commerce. There are a number of issues with global development of web based commerce. The most important issues are evaluation methods and cultural differences among users. Over the years a number of modular studies have been done at the Department of Industrial Engineering, University of Nebraska. This presents an overview of some of the studies that have some ramification for web usability.

2 Comparison of Heuristics Evaluation and Usability Testing

Usability plays a crucial role as user experience is emphasized above anything else. Thus, it is essential to create a well-designed Website that is highly usable. The question is one of deciding what constitutes a well-designed site and how to evaluate the same? Different usability evaluation techniques have been developed and incorporated

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 243–249, 2007.
© Springer-Verlag Berlin Heidelberg 2007

into the design and development of Websites. Among these techniques user testing and heuristic analysis are perhaps two of the most popular ones.

The main objectives of this study were:

1. Evaluate four commercial websites with both user testing and heuristic analysis,
2. Compare the efficiency and effectiveness of both the methods

A total of four commercial websites were evaluated for this study, the first two websites were considered to have average number of usability problems while the other two were considered to have high number of usability problems. Bad websites represented the user interface in the early stages of the development process that have abundant amount of usability problems. On the other hand, good websites represented the user interface in the later stage of the development process with lesser usability problems compared to the bad websites.

A total of 5 scenarios that represented typical site usage situations that might be encountered in real life were given to both the users and evaluators. The scenarios for both user testing and heuristic analysis were essentially the same but varied in terms of the degree of detail. The scenarios for user testing were more detailed while the scenarios for heuristic analysis were more open ended. A total of 12 users for user testing, and 9 evaluators for heuristic analysis participated in this study.

The analysis revealed several important findings as detailed below:

1. First of all, the proportions of common problems were small for both good and
2. bad websites, which were approximately 7% and 11%,
3. Second of all, proportion of the common problems found increased from good to
4. bad websites, about 4%.
5. Third of all, heuristic analysis found the most problems in both types of websites,
6. approximately 58%-61%.
7. Fourth of all, heuristic analysis is equally effective in both good and bad websites
8. environment.
9. Finally, user testing found less problems in bad websites compared to good
10. websites. These conclusions were consistent throughout all four websites.

The main premise for this study was that both user testing and heuristic analysis should be used as both an evaluation tool and as a method for guiding design improvements. The study confirmed the premise. The results indicated that both methods are needed as both methods find very few common problems. It is known that both user testing and heuristic analysis are based on very different fundamentals. User testing relies mainly on the experience and comments of the users and is usually conducted in a scenario-based environment. As a result user testing would usually evaluate according to what already exists, rather than to what is possible. On the other hand heuristic analysis relies mainly on the expertise and knowledge of human-factors engineer that would evaluate the web site based on a set off heuristics. Therefore both methods find different types of problems.

In summary, both user testing and heuristic analysis are needed in a usability study. In order to reap the optimal benefits, both user testing and heuristic analysis should, preferably, be used in different stages of the user interface design process.

3 Comparison of Remote Testing Versus Direct Testing

Usability testing continues to be the primary method of evaluation of web sites. Though time consuming and resource intensive, they are the most predominantly used. Usability studies are typically performed under direct supervision of the experimenter, under a controlled environment, which often poses a constraint. Further, the down side of running the evaluation in a controlled environment is that it does not recreate the Internet browsing environment from the subject's own computer. A good way to solve this problem is to have the subject evaluate the website on his/her own computer and in his/her own environment. Scenarios are mailed to the subject and the evaluator can monitor the evaluation process from his/her lab through remote access. *The working hypotheses was that there would be no differences between direct and remote usability.*

Two groups of student subjects (Asians and Americans) participated in this experiment. Two types of usability testing (direct and remote) were performed on two sites. In the remote testing, the subjects were in their own room separated from the moderator. The subjects answered the user profile questionnaire, evaluated the websites through 5 scenarios, and did a post evaluation questionnaire. As the subjects were doing the evaluation, PCAnywhere captured the subjects' computer screens in the form of video and the video was saved in the moderator's computer. In the direct testing, the evaluation took place at the Department of Industrial Engineering Ergonomics Lab. The subjects and the moderator were physically in the same room. The procedure was identical to the remote testing procedure with one difference, when the subjects got stuck in any of those scenarios, the moderator was physically there to give guidance to the subjects.

All of the results and analyses in this study suggested that remote usability testing provided the same result as traditional usability testing regardless of ethnicity. Usability testing has become one of the important tools in evaluating a website and many engineers are looking at ways to cut costs without compromising the effectiveness of the method. By using the method of remotely conducting the usability testing, costs can be cut down tremendously. There was no interaction between sites and method, and between ethnicity and method, suggesting the robustness of this method.

4 Cultural Differences in Web Usability

As per the name, World Wide Web has evolved as a medium for international communication, participation and transaction for a multi-cultured environment. Usability studies in past have been mainly limited to language impact on global market. Site visitors must be able to navigate freely, confidently and comfortably through a site in order to find, enjoy and make use of its contents. This requires the web designers to consider the cultural background and behavioral pattern of the users.

Culture is a learned phenomenon, which derives from one's social environment. It is the collective programming of the mind which distinguishes the members of one group or category of people from another. A dimension is an aspect of a culture that can be measured relative to another culture. Geert Hofstede in his classic study of cultures proposes that each national culture has five major dimensions which can be

its identification mark. They are Power Distance, Collectivism vs. individualism, Femininity vs. masculinity, Uncertainty avoidance and Long vs. short term orientation.

The objectives of this study revolved around cultural dimensions effecting web usability.

1. To measure the quantitative attributes for different cultures based on their cultural dimension scores.
2. To find whether the web performance of people differs based on their cultural background in terms of time taken to accomplish the given task and number of pages visited during the task.
3. To find whether people have different web preferences in accordance with their cultural dimensions

To measure the cultural dimensions, Value Survey Model questionnaire, developed by Geert Hofstede, was used. The questionnaire contains 25 questions which measures 5 dimensions for each culture. The experiment contained total of 4 tests to be performed by each of 20 participants from 3 nationalities. Subjects were first asked to solve the Value Survey Model (VSM) questionnaire. It was followed by Web Stereotype questionnaire. Then they were asked to surf the sites identified for the study and to complete a task following the simple guidelines. This task was monitored by Web Logger software. Finally a card sorting test was performed to determine ethnic differences, if any, in developing information architecture.

This study was based on the hypothesis that culture does have an important effect on web usability. The quantitative analysis was carried out to find the significance in total task time, number of pages visited and ethnicity. The descriptive analysis was performed to establish the relation between various cultural dimensions and user's web preferences. By analyzing VSM scores for each ethnicity and by using Marcus guidelines, we could interpret the web behavior for each country. The regression results between task time, number of clicks and ethnicity were quite promising. The ethnicity showed significant relation with total task time and number of clicks. Results concluded that ethnicity does have an important role to consider in web usability. Web surfing performance has shown significant difference in total task time according to user's cultural background. But the study has failed to show a distinct difference in web stereotype questionnaire and so user's web preferences. It also lacks some more research on card sorting exercise where it is unable to indicate the subtle differences in spatial arrangement of the information. As many users are not the web designers, they tend to arrange the card sorting data according to their impression of the other popular sites.

In summary, this study has contributed towards progress of web usability field by adding the information that cultural dimensions should be considered while designing the web sites for international users to capture global attention and increasing the profitability of the business.

5 Web Personalization Study

Web personalization can be described, as any action that makes the Web experience of a user personalized to the user's taste. This experience can be something as casual as browsing the Web or as economically significant as trading stocks or leasing an

apartment. User satisfaction is the ultimate aim of personalization. The primary objective of this study was to determine if user's surfing behavior could be predicted through their cognitive style. Stated formally, the objectives were:

Hypothesis I: Cognitive styles affect web surfing styles
Hypothesis II: Presence of discrete relation between individual cognitive styles and web surfing behavior.

A number of methodological details had to be sorted out before data collection. They are:

1. A general web surfing task had to be designed as to capture all complexities of web surfing behavior,
2. A method of capturing objective surfing data had to be designed, and
3. A battery of independent tasks that would be a valid predictor of surfing behavior had to be designed

A task was to be designed with following features in its way of execution. It should give subject sufficient sense of control with reasonable complexity so as to capture reasonable amount of data and there should be no time limit which might limit subjects natural response to task. Task was designed within Cognitive design model prescribed by Norman (1991). This study wanted to compare cognitive spans with cognitive processing in web surfing, which needs web-browsing data to be in comparable format to span scores from psychometric tests. Weblogger software developed by XEROX PARC was found with such capability. Cognitive spans were selected as mental measures for individual differences. Three spans, i.e., counting spans, operation span and reading span were used. Individual differences that are expected in these dimensions of working memory capacity primarily reflect differences in capability for controlled processing. Total 50 subjects participated in the experiment.

The analysis was based on four variables: three span scores and time. At first, descriptive analysis was carried out to check the suitability of data. Regression of span scores against total task time highlighted their interaction. The next phase of analysis was divided into task study, between page analysis and within page analysis. Task study consisted of analysis to develop predictive relation cognitive spans and total task time. Between pages analysis studied the individual web page and its interaction with neighboring pages in URL maps. Time span for each web page was analyzed with mean time analysis across the subjects' cognitive span scores. Finally, within page analysis was consisted of analysis of subject's performance for individual web page.

This study was carried out with intention of finding predictive capacity of cognitive style of individuals for their web surfing behavior. It was based on hypothesis that website surfing is dependent on users cognitive span (attention span). This study had more interest in finding whether the user coming to site falls in one of the groups of spans, which can help predict further course for such user, basically adapting the site to the individual. Results concluded cognitive span do predict surfing behavior of individual to some extent. For high cognitive ability individuals, web surfing performance was very much improved. This study failed to clearly define low cognition individuals who had their performance overlapped by medium span individuals. Psychometric tests were used to measure and quantify this natural performance in web

surfing. But the tests failed to capture the natural performance in there scores. Psychometric testing needs improvement in its scope to capture human cognition free of outside influences. Personalization can go to much higher extent in helping the user navigate efficiently and achieve the goal. But the designer still needs to answer whether in spite of all the efforts, will it be economically feasible to retain and maintain personalization soft wares and tools?

6 Discussion

What do all these mean for a web designer and web user? Web has come to stay. Ever since real time interaction on the web became a reality, growth in World Wide Web has been mind boggling. It has rendered commerce and all aspects of life truly global. With regards to web, there are four perspectives: that of a user, designer, developer and main organization which wants web presence. From a user view point two issues stand out: a) the experience of using the web (a particular site for example) has to be a pleasant one, and b) the interaction (transaction) has to be natural. The web site, for the organization seeking web presence, has to be a) simple, b) exhaustive, c) ensure customer retention and d) provide a pleasurable user experience. The designer has to have a good process in place for designing the site. The designers should always remember that competition is just one click away. Finally for the developer and the maintainer the web has to be easy to maintain and robust. Our studies, those summarized here and others, deal more with designers' and users' perspectives.

It is clear the both heuristic evaluation and usability testing are valid testing. They are *complementary* methods of evaluation and not *competing* methods. They identify different sets of problems, all of which have to be corrected before the final release of the web site. Both methods are equally efficient and effective in addressing different categories of usability problems.

It is also apparent that there are no differences between remote usability testing and traditional usability testing. This has two ramifications, reduced cost and better fidelity. The site can be tested with actual potential users rather than with trained usability subjects. From the next study reported here, it is clear that significant cultural differences exist in the manner the subjects perform the tasks on the websites. The results also indicated that cultural dimensions have significant effects on user's web preferences. Web designers need to consider the cultural background of the target users while designing the websites. Finally, the study on personalization suggests cognitive span scores are not strong enough to form association rule with individual difference clusters of web surfing behavior.

Bibliography

1. Deshmulkh, N.: Personalization of Webs sites, A Master's theses, Department of Industrial Engineering, University of Nebraska (2002)
2. Dixit, A.: Cultural Differences in web usability, A Master's theses, Department of Industrial Engineering, University of Nebraska (2003)
3. Liew, W.-C.: UsabilityTesting: Remote Vs. Direct. A Master's theses, Department of Industrial Engineering, University of Nebraska (2002)

4. Tan, W.S.: Comparison of heuristic evaluation and usability testing, A Master's theses, Department of Industrial Engineering, University of Nebraska (2002)
5. Deshmuk, N., Dixit, A., Bishu, R.R.: Web Personalization: Study Of Effect of Cognitive Style. In: The Proceedings of the Fourth International Conference on Work with Computing System, pp. 553–557 Kualalampur (2004)
6. Dixit, A., Bishu, R.R.: Cultural Differences in Web Usability. In: The Proceedings of IEA Congress, Seol, Korea. (Page numbers not provided) (2003)
7. Deshmukh, N.B, Dixit, A, Bishu, R.R.: Web personalization: Study of effect of cognitive style on web surfing, Presented at the 48th Annual Meeting of the Human Factors and Ergonomic Society, Denver (2003)
8. Liew, W.-C., Bishu, R.: Web Usability: Remote versus Direct Testing. In: The Proceedings of IEA Congress, Seol, Korea. (Page numbers not provided) (2003)
9. Tan, W.S., Bishu, R.R.: Which Is a Better Method of Web Evaluation? A Comparison of User Testing and Heuristic Evaluation. In: The Proceedings of the 46th Annual Conference of Human Factors and Ergonomics Society, pp. 1256–1260, Baltimore (2002)

The Impact of Different Icon Sets on the Usability of a Word Processor

T.R. Beelders, P.J. Blignaut, T. McDonald, and E. Dednam

Department of Computer Science and Informatics, University of the Free State, PO Box 339, Bloemfontein, South Africa, 9300

Abstract. This paper discusses the results of usability tests obtained when testing different sets of icons in a word processor environment. An alternative set of icons was developed for a subset of word processor functions and compared to the standard icons. The score obtained for completed tasks as well as the time taken to complete tasks successfully were evaluated. Results indicate that the score is not affected by the icons used in the interface. It was noted that word processor expertise and the icons used have a significant effect on the time taken to complete some tasks. However, each of these factors exhibits an effect in only a single task completed in the prototype. Possible reasons for the significant difference are discussed.

Keywords: Usability, icons, interface.

1 Introduction

The advent of the graphical user interface (GUI) resulted in an increase in the use of icons within computer applications [1]. Users have exhibited distinguishable preferences for interface components such as language, navigation, symbols and colour use [2]. These facts motivate the need for careful consideration of, amongst others, translation and icon development in user interfaces [3] – factors which could have an impact on product usability.

This paper will discuss some of the available literature on icons and usability. An outline of the research methodology that was used will be given, followed by a detailed discussion of the experiment results. Finally, a conclusion, based on the analysis of the results, will be drawn.

1.1 Usability

According to the International Standards Organisation (ISO) standard 9126-1 usability is "the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions". This definition is further expanded upon in ISO 9241-11 where usability is defined as "the 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" [4].

Using the above definitions, four distinct components of usability can be identified, namely effectiveness, efficiency, satisfaction and learnability. These components are defined as follows:

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 250–257, 2007.
© Springer-Verlag Berlin Heidelberg 2007

- Effectiveness is how well the user is able to achieve that which must be done by using the system [4] and can be measured in terms of accuracy and completeness [5].
- Efficiency is the amount of resources required to complete the desired task [4].
- Satisfaction is the subjective feeling the user has about using the system [4].
- Learnability measures not only the time taken for a user to become familiarised with the system but also how well the user is able to remember system functionality [5].

Shneiderman [6] lists a set of five measurable objectives that can be measured in order to determine the usability of a product. These measurements of time to learn, speed of performance, rate of errors by users, retention over time and subjective satisfaction allow for specific and controlled evaluation of a software application [6]. The available usability models also provide a number of measurements which can be used by developers to comprehensively test the usability of a product [7].

1.2 Icons

Icons are a common interface component that employ images to represent an object or an action that can be carried out by the user [8]. Continued use of icons has been attributed to the fact that they are easier for users to learn [1], [9] and to use [1]. Their use also increases the productivity of the user since recognition is generally faster for a picture than for text [1], [8].

One disadvantage of icons is that they can easily be misinterpreted by users if the chosen image invokes unintended associations [8] – the picture that "speaks a thousand words may say a thousand different words to different viewers" [10].

Zammit [10] investigated the effectiveness of pictorial and text icons with a group of 11 and 12 year olds and found that neither the pictorial nor the text icons were always immediately recognisable to the users. Users' accuracy has been shown to be the highest when selecting from a mixed format of text and graphics in a menu structure as opposed to a graphics only or text only menu structure [1]. However, no discernable difference in the time taken to make a selection was detected between the three formats [1].

This research undertook to test the usability of a set of preferred icons chosen by non-computer literate users by means of empirical testing. To complete the set of chosen functions, the remainder of the icons were developed by means of a brainstorming session. By comparing the alternative set of icons to the standard word processor icons, it can be determined whether the alternatives are better suited to South African users and in so doing, establish whether or not there is a need to develop new word processor icons for a South African audience.

2 Research Procedure

2.1 Method

By making use of a simple word processor application the effect of different sets of icons on the usability of a product was tested. Two sets of abstract pictorial icons [10] were used in order to determine if the icons used influence product usability.

2.2 Word Processor Prototype

A small word processor application was developed in order to test the subjects. The word processor possessed minimal capabilities, while still ensuring that it was representative of a fully-fledged word processor or advanced text editor. Functions which were incorporated into the word processor prototype included document handling (e.g. open and close), text formatting (e.g. font size and style) and text manipulation (e.g. copy, cut and paste).

The prototype also allowed for capturing of the users' demographic information, such as age, gender and language.

Users were required to complete a number of small tasks, representative of common word processor tasks. The tasks were displayed sequentially and individually at the bottom of the word processor window (Fig. 1) and could be completed solely by making use of either a toolbar shortcut (icon) or a menu option.

The prototype allowed for real-time evaluation of the tasks as the user completed each one. A number of measurements were also captured for each task. These include the number of menu options and toolbar shortcuts selected by the user, together with the number of keystrokes and mouse clicks and the time required to complete the task.

2.3 Interface

Two sets of icons were used in the different interfaces, namely the standard icons currently found in the Microsoft Office packages and an alternative set of icons obtained from previous studies [11] and via two brainstorming sessions (Fig. 1).

The set of icons obtained during the first brainstorming session were distributed amongst potential word processor users. Given the complete set of icons, respondents were required to indicate which icon they would choose for each of a number of listed word processor functions. All the icons were available to be chosen for each function and icons could be chosen more than once. The alternative icons used in this study for the functions of **Open**, **Close**, **Save**, **Cut**, **Copy** and **Paste**, were chosen as the preferred icon by the group of non-computer literate respondents. There were two icons chosen by the same number of respondents for the **Close** function. Therefore, the icon for **Close** was selected by a process of elimination since one of the icons was also chosen as the preferred **Save** function by a large margin. A number of these icons were confirmed in the same manner in an independent study undertaken by Teklebrhan and Blignaut [11].

The remainder of the icons were developed during a second brainstorming session and included in the design without confirmation by non-computer literate users. The icons were developed to provide more context for novice and first-time users. For example, the icons used for **Bold**, **Italic** and **Underline** consisted of a bold, italic or underlined capital letter "F" respectively. This was done in an effort to convey to the user the font changes that would occur if the function were invoked. By using the same letter throughout and placing them adjacent to one another on the toolbar, it allows for easier visualisation of the font styling (Fig. 2). It was hoped that by developing in such a manner that novice and first-time users would easily relate to the concepts depicted by the icons.

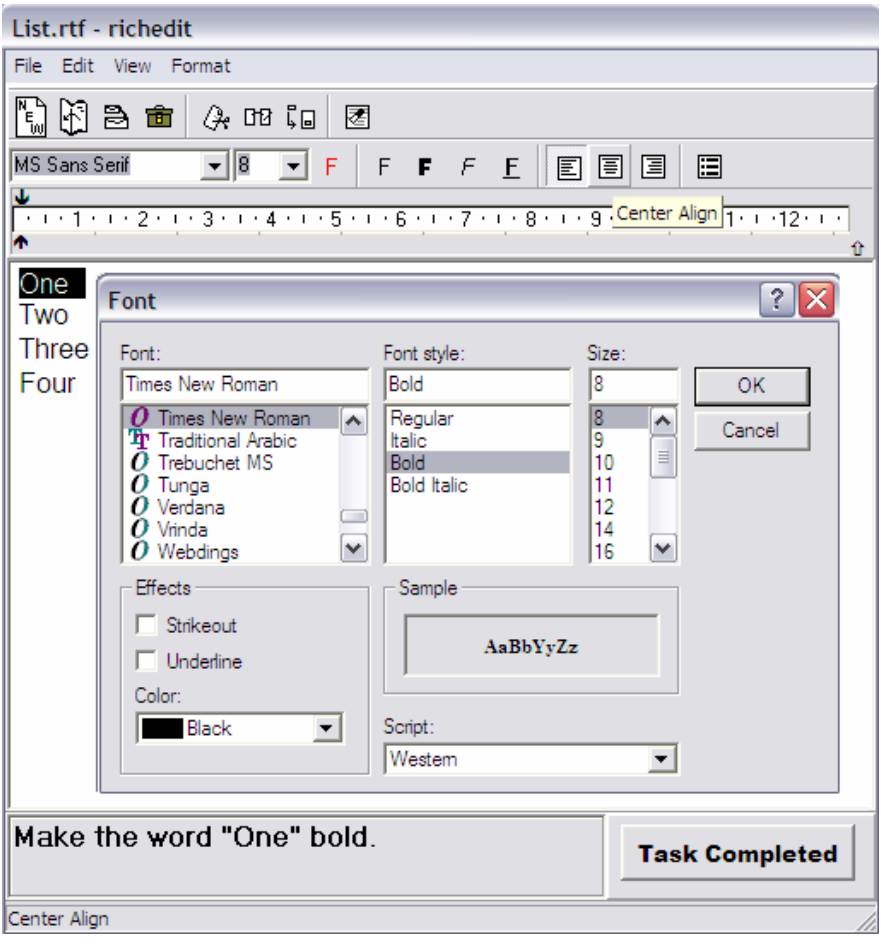


Fig. 1. Word processor prototype with alternative icons

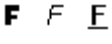


Fig. 2. Font styling icons

To ensure that the effect of the icons could be tested without interference from other interface components, the interface had neither menus nor tooltips. The user had to rely entirely on interpretation of the icon.

In summary, the two interfaces tested were:

- Standard icons with no menu and no tooltips.
- Alternative icons with no menu and no tooltips.

2.4 Subjects

The test subjects consisted of first year university students that were taking a basic computer literacy course. Test subjects spoke a variety of languages, including English, Afrikaans, Sesotho and isiZulu. All subjects were conversant in either English or Afrikaans as these are the tuition languages of the university. The participants provided for different levels of word processor expertise. There were 61 female and 37 male participants who completed the test on either a standard interface or an alternative interface.

2.5 Testing Environment

The test was conducted during the first practical session of the course. This was before the users had received any instruction in word processor packages. Up until that point they had only been taught basic Windows usage.

Each participant was randomly assigned to one of the interface groups. After all the practical sessions had been completed there were 98 tests that had been completed, of which 47 were completed on the interface with standard icons and 51 using the alternative icon interface.

3 Analysis

The usability measures that were analysed were (a) an overall score (discussed below), (b) time taken to complete a task, (c) number of actions required to complete a task, (d) number of errors incurred whilst completing a task, (e) user satisfaction. The number of correct and incorrect answers was also compared for each task. Of these only the effectiveness measurement of score and the efficiency measurement of time [7] will be discussed in this paper.

3.1 Independent Variables

The independent variables used in the analysis were the interface employed during the test and the word processor expertise of the user.

Each user was classified as either a first time or an expert [6] word processor user based on their level of experience with a word processor application, together with the frequency with which they make use of such an application.

3.2 Dependent Variables

The two dependent variables discussed in this paper are the score for each user and the time taken to complete each task. To calculate the score each task was assigned a difficulty index based on the minimum number of actions and inferences required to complete the task successfully. This allowed for a weighted score to be computed for each user. The time taken to complete each task was measured in seconds and then converted to 1/time for further analysis.

3.3 Analysis of Score

The evaluation of the score was done by means of a 2 x 2 between subjects factorial ANOVA. The following hypotheses were formulated for the score:

1. $H_{0,1}$: Word processor expertise has no effect on the score achieved .
2. $H_{0,2}$: The interface used has no effect on the score achieved.

The word processor expertise of the user had no effect on the achieved score since $H_{0,1}$ could not be rejected ($F_{Expertise}(1, 94)= 0.989$, $p = 0.322$). $H_{0,2}$ could not be rejected ($F_{Interface}(1, 94)= 1.192$, $p = 0.278$), leading to the conclusion that the interface used during the test did not have a significant effect on the achieved score of the user.

3.4 Analysis of Time

The time was evaluated individually for each task by means of a 2 x 2 between subjects factorial ANOVA. Only those tasks that were completed successfully were included in the analysis [12]. The following hypotheses were formulated for the time variable:

1. $H_{0,1}$: Word processor expertise has no effect on the time taken to complete a task successfully .
2. $H_{0,2:}$ The interface used has no effect on the time taken to complete a task successfully.

$H_{0,1}$ could be rejected for only one task ($F_{Expertise}(1, 81) = 4.302$, $p < 0.05$), where expert users performed significantly better than first time users. The task required users to change the font colour of a word. Two possible explanations could be offered for this difference. Firstly, it was observed during the test that many users experienced difficulty in grasping the concept that the drop-down box containing the font colour can be expanded to reveal a wider selection of colours. Secondly, after changing the font colour, the selection distorts the actual colour of the word. For example, green coloured font appears to be purple when selected. This phenomenon confused users not familiar with the effects that highlighting has on the font appearance. These two observations could possibly have caused some hesitation and confusion on the part of first time users, thus leading to a longer completion time for these users. Evaluation of the number of actions required and the number of errors incurred during completion of the task could provide more information on the cause of the difference.

A second task, which appears slightly later in the test, required users to change the colour of a whole sentence. There was no significant difference exhibited between the users for this second task. This seems to indicate that users retained the knowledge obtained in the previous task and did not experience the same problems again.

Users of the standard icons performed significantly better on the task that required users to close the text document ($F_{Interface}(1, 45) = 9.797$, $p < 0.05$), allowing $H_{0,2}$ to be rejected for that task. The alternative icon for the **Close** function was chosen by questionnaire respondents, but the results of this task show that it did not communicate the concept of **Close** as clearly as the standard icon. In fact, the icon chosen by the respondents was actually designed as an alternative for an electronic mail interface. Taking into consideration that choices of non-computer literate users were split evenly between the icon eventually used for **Save** and the one used for **Close**, it may be

pointed out that perhaps the entire concept of closing a document needs to be explained more clearly to novice or first time users.

To place these icons in perspective, they are shown below in Table 1.

The obtained result indicates that although users show a preference for a certain icon, it does not necessarily improve the usability of the product. Icons that are used should be chosen with care and developers should ensure that the icon does indeed convey the intended meaning or concept.

Table 1. Standard and alternative icons

	Standard	Alternative
Save		
Close		

4 Conclusion

The interface had very little effect on the usability of the word processor, a finding which corroborates those of Kacmar and Carey [1] where time is concerned. The only significant difference between the users of the different interface occurred when using an icon that potentially did not convey the meaning of the function clearly to the user. This supports the assertion that careful consideration should be given to the development of icons [3]. The fact that the icon in question was chosen as the preferential icon by questionnaire respondents could indicate a distinct lack of understanding for the concept portrayed by the icon.

Word processor experience only had a significant effect on the task that required use of a complex dialog box, a situation in which it is understandable that a first-time user would show some hesitancy or uncertainty. Subsequent tasks using the same dialog box showed no significant performance difference between users – an indication that users do retain the learned concepts, at least for a short period of time. It would be interesting to test whether users are able to retain this knowledge over a longer period of time than simply between two tasks.

Results would indicate that there is no need for development of an alternate set of icons for South African users. Standard icons appear to be intuitive enough that they correctly convey that which they attempt to represent. Rather, proper explanation of word processing concepts and functions is needed. Given enough time and practice, it appears that most users will be able to master the usage of a word processor application.

References

1. Kacmar, C.J., Carey, J.M.: Assessing the usability of icons in user interfaces. Behaviour and Information Technology 10(6), 443–457 (1991)
2. Cyr, D.,Trevor-Smith, H.: Localization of web design: An empirical comparison of German, Japanese, and U.S. website characteristics. Retrieved 24 February 2005 from www.eloyalty.ca/docs/Localization_of_Web_Design.pdf

3. Johns, S.M.: Colors, buttons, words and culture: Designing software for the global community. Retrieved 24 February 2005 from http://www.sensi.org/~alec/locale/other/colore~1.htm
4. ISO9241. ISO 9241-11: Ergonomic requirements for office work with visual display terminals. Beuth, Berlin (1998)
5. Cato, J.: User-centered web design. Addison-Wesley, Great Britain (2001)
6. Shneiderman, B.: Designing the user interface: Strategies for effective human-computer interaction, 3rd edn. Addison-Wesley, United States of America (1998)
7. Abran, A., Khelfi, A., Surya, W., Suryn, W., Seffah, A.: Usability meanings and interpretations in ISO standards. Software Quality Journal 11, 325–338 (2003)
8. Benbasat, I., Todd, P.: An experimental investigation of interface design alternatives: icon vs. text and direct manipulation vs. menus. International Journal of Man-Machine Studies 38, 369–402 (1993)
9. Schild, W., Power, L.R., Karnaugh, M.: Pictureworld: A concept for future office system. IBM, Thomas J. Watson Research Center, Yorktown Heights, NY, RC 8384 (#36518) (1980)
10. Zammit, K.: Computer icons: a picture says a thousand words or does it? Journal of Educational Computing Research 23(2), 217–231 (2000)
11. Teklebrhan, R., Blignaut, P.: A study on the effect of Western designed metaphors in some culture groups in South Africa. Department of Computer Science and Informatics: University of the Free State, Technical Report 2005/02 (2005)
12. Nielsen, J.: Outliers and luck in user performance. Alertbox (March 6, 2006) http://www.useit.com/alertbox/outlier_performance.html

Systems Development Methods and Usability in Norway: An Industrial Perspective

Bendik Bygstad[1], Gheorghita Ghinea[1,2], and Eivind Brevik[1]

[1] Norwegian School of Information Technology, Oslo, Norway
[2] School of Information Systems and Computing, Brunel University, UK
bygen@nith.no, george.ghinea@brunel.ac.uk, eivind.brevik@nith.no

Abstract. This paper investigates the relationship between traditional systems development methodologies and usability, through a survey of 78 Norwegian IT companies. Building on previous research we proposed two hypotheses; (1) that software companies will generally pay lip service to usability, but do not prioritize it in industrial projects, and (2) that systems development methods and usability are perceived as not being integrated. We find support for both hypotheses. Thus, the use of systems development methods is fairly stable, confirming earlier research. Most companies do not use a formal method, and of those who do, the majority use their own method. Generally, the use of methods is rather pragmatic: Companies that do not use formal methods report that they use elements from such methods. Further, companies that use their own method import elements from standardised methods into their own.

1 Introduction

This paper investigates the relationship between two important disciplines of modern systems development; the use of systems development methods and the concepts and techniques of usability.

Systems development methods (SDM) have been in use the past forty years and constitute a core part of modern software engineering. Still, they represent a thorny issue, both because their effectiveness has been challenged [6], [15] and because of the continuous wars between proponents of different methods [10]. During the 1990s most methods became iterative and incremental, acknowledging the emergent nature of software development. Well-known examples are Rational Unified Process [8], DSDM [14], Microsoft Solutions Framework [11] and XP [1].

Usability, on the other hand, emerged during the late 1980s, and was embraced in the 1990s by parts of the software industry as a response to the challenges that web based software put on developers. The body of knowledge of usability is large and includes various perspectives, from usability engineering [12] to more context-oriented approaches [2].

This paper investigates empirically, through a survey among Norwegian IT companies, the relationship between SDM and usability in current industry practice. We investigate which SDMs that are adopted, and to which degree the companies have

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 258–266, 2007.
© Springer-Verlag Berlin Heidelberg 2007

adopted usability techniques. These findings are used to investigate our core assumption – that systems development and usability are both accepted as best practices in principle, but not yet integrated in a full process.

The paper is structured as follows. In section 2 we discuss findings in earlier research, and present our two hypotheses. Then, in section 3, we briefly present our research method. In section 4 the result of research will be presented followed by a discussion. Section 5 concludes and points to further research.

2 Assumptions and Hypotheses

Although SDMs and usability have some similarities (they are both applied disciplines, and they play important roles in systems development) their differences are much more obvious. While SDMs originated from systems engineering and software economics [13] in the late 1960s, usability was developed in the late 1980s and early 1990s from HCI research, cognitive psychology and phenomenology. While systems development was – with some notable exceptions - mainly concerned about the inner workings of the system, usability focused on the user. Thus, the role of the user is different; in systems development the user is a means to elicit requirements [8], while for usability work the users are the prime means for designing the system [7], [12].

Systems development theorists tend to play down these differences, arguing that usability may easily be integrated into the formal frameworks [9]. Oppositely, usability researchers have argued that these differences add up to two different cultures of systems development, and have called for new approaches to counter the basically technical approach of SDMs. For example, Boivie et al [3] concluded – after a review of this relationship – somewhat pessimistically:

> "*We believe that one of the main difficulties with incorporating User Centric Systems Development in existing processes is that it requires a great deal more than simply adding a few activities to existing processes. It requires new development approaches, new methods, new roles, new ways of planning and allocating resources etc. Moreover, a user-centered approach changes the relationship between the user/client organization and the development organization (..)*".

Our point of departure is that these issues should be investigated in an industrial context. From this discussion we propose two hypotheses. The first is concerned with the general status of usability in systems development.

H1: Software companies will generally pay lip service to usability, but do not prioritize it in industrial projects.

This hypothesis assumes that there is a gap between intention and reality; that the companies will express concern for usability, but not be willing to use resources on it in industrial projects with strong time and cost pressures.

The second hypothesis is concerned with the perceived relationship between systems development methods and usability. We assume that most companies use some kind of method and that they also relate to usability issues. However, we do not believe these are integrated in the practices of the development projects.

H2: Systems development methods and usability are perceived by practitioners as not being integrated

In the next section we outline how the hypotheses were investigated.

3 Research Method

This section will first give a description of the sampling and sampling design that has been used. Then research design and analysis of survey responses are determined.

3.1 Sampling and Sampling Design

The greatest sampling challenge in this type of research is to identify which companies that actually engage in systems development [6]. This study builds on similar studies done in Norway in 2002, 2003 and 2004 [4], [5], where a great deal of effort was put into establishing a population of Norwegian IT companies that engage in systems development. Ideally, all companies involved in software development in Norway should be defined as the population for this research. This includes general private companies and public organisations as well as professional companies within the IT sector.

Earlier studies showed, however, that response rates from general private and public companies were too low to be useful. Thus, the population was limited to IT companies in the following three different Norwegian industrial classification (IC) codes:

7220000 System- and software consulting
7260001 IT consulting
7260003 IT services

Our sample was collected from two sources. First it consists of the 194 companies that accepted to participate in 2003. Second, this was supplemented by 65 companies that participate as partners in NITH student development projects, which we knew were engaged in systems development. Of course, this sampling strategy puts some limitations on the implications of our findings, which we will return to in our discussion.

3.2 Research Design

A questionnaire was designed, with 5 questions on SDMs and 8 questions on usability. We also asked how many persons were engaged in systems development in the company.

The survey was implemented electronically by using the QuestBack system[1]. This system is based on e-mail distribution of a link to the actual survey and replies via a web browser on the Internet. The QuestBack system has an automatic reminder, which was scheduled once to those who had not responded after the request to participate in the survey was sent out. After about a four weeks' period, the survey was closed with 87 responses, representing a response rate of 33%.

[1] www.questback.com

4 Results and Discussion

This section presents the results and discussion, and is divided into three parts: (1) Adoption of SDM (2) Usability in requirements and testing and (3) The relationship between SDMs and usability. The first two sections are descriptive, while we test our hypotheses in part 3.

4.1 Adoption of SDM

Respondents were asked whether or not they were using a formal SDM.

Table 1. Formal SDM use

Answer	N	Percent
Yes	27	35 %
We do not use a formal SDM, but we use a number of techniques and tools	45	57 %
No	6	8 %
SUM	78	100 %

As shown in Table 1 the majority do not use a formal method, but a number of techniques and tools. Respondents that answered 'yes' were then asked to indicate which formal SDMs that were in use. The result is shown in Table 2.

Table 2. Breakdown of formal SDMs used in Norwegian companies

Method	Use 2006	Use 2003	Use 2002
Own method	**68%**	78 %	79 %
RUP	**29%**	29 %	23 %
XP/Agile methods	**18%**	21 %	17 %
MSF	**29%**	19 %	21 %
OPEN	**0%**	11 %	0 %
PSO	**0%**	7 %	21 %
Other methods	**19%**	10 %	13 %

The sum of percentages is greater than 100 % because some companies use more than one method. A large majority, 68 %, of software development companies uses their own method. This is in line with the findings for the 2003 and 2002 survey. The numbers do not provide evidence of a significant change in the usage of commercial methods. Rather, they suggest that companies tend to stick to a certain method, and are reluctant to change. The comments from the companies illustrate this point; they are generally quite satisfied with their choice of method.

4.2 Adoption of Usability Techniques

Designing for usability typically involves establishing user requirements for a new system, iterative design and testing with representative users. Thus, in order to examine the interplay between usability and system development methods, in our survey we specifically sought to explore to which degree usability was included in the system requirements and the degree of usability testing. Usability in requirements was measured by two questions, the first being "When will you include usability in requirements?" The result is shown in Table 3.

Table 3. Usability and requirements

Answer	N	Percent
Always	55	72 %
Only if usability problems emerge during the project	8	10 %
Only if the customer demands it	12	15 %
Only if we have an internal usability specialist available	2	3 %
Sum	**77**	**100 %**

The second question was "*How do you collect requirements for usability?*" Results are shown in Table 4.

Table 4. Usability and requirements (multiple answers possible)

Answer	Percent
Interviewing users	67 %
Best practice from earlier projects	71 %
Books, Internet resources	19 %
Other	12 %

Respondents were also asked two questions on usability testing. The first was "*How many users are typically engaged in usability testing?*" As Table 5 shows, the samples of users in testing are generally small, most being less than 10 users.

Table 5. Number of users involved in usability testing

Answer	Percent
1-10 users	66 %
11-50 users	21 %
More than 50 users	3 %
We do not test usability	10 %
Sum	**100 %**

Table 6 show how these users were selected. It shows that 40% of the respondents report that they select a representative sample of users.

Table 6. Selection criteria for users in usability testing

Answer	Percent
Arbitrary sample of users	5 %
Representative sample of users	40 %
Own employees	9 %
Customer's employees	23 %
Other	15 %
Do not test usability	8 %
Sum	**100 %**

Summarizing the findings on usability the results shows that the majority of the respondents include usability in their requirements, and that they also collect usability requirements by including users in the process (Table 3 and Table 4). In usability testing, however, the number of users seems quite small, as most of the companies only include less than 10 users (Table 5). Furthermore, only about 40 % of the users selected for testing are a representative sample of the users.

4.3 The Relationship Between SDM and Usability

Returning to our two hypotheses we first assumed:

- H1: Software companies will generally pay lip service to usability, but not prioritize it in industrial projects.

To investigate this hypothesis we first assess the answers of two general questions on usability. The respondents were asked – in general terms - how important usability requirements and usability testing was for the success of their projects. The result is showed in Table 7.

Table 7. Usability requirements, usability testing - and project success

Answer	Usability requirements	Usability Testing
6- Very important	33 %	14 %
5	38 %	23 %
4	21 %	31 %
3	6 %	19 %
2	1 %	6 %
1 – Quite unimportant	1 %	5 %
Sum	**100 %**	**100 %**

As Table 7 shows the majority of the companies thinks usability is important for the success of their projects. Somewhat surprising, usability requirements are considered more important than usability testing.

However, when assessing the answers of the more concrete questions on usability activities in projects, the results show a different picture. Concerning usability requirements, 72 % of the companies always include it, and almost 67 % also interview the users, as showed in Table 4 and Table 5. On the other hand, only 40 % of the companies use a representative sample of users for usability testing (Table 6). Further, the number of users engaged in usability testing is generally quite small, as showed in Table 5.

In concluding, we find that our first hypothesis is supported by our empirical materials. There is a gap between intention and reality: the companies express interest and concern for usability, but this stance is not corroborated by their subsequent responses, which reveal that they are less willing to use resources on it in industrial projects with strong time and cost pressures.

Our second hypothesis was:

- H2: Systems development methods and usability are perceived by practitioners as not being integrated

The respondents were asked "*To which degree do you think that usability is integrated in your systems development method (whether you use a formal SDM or not)?*" The result is shown below in Table 8.

Table 8. To which degree is usability integrated in systems development method?

Answer	N	Percent
6- To a large degree	11	14 %
5	18	23 %
4	21	26 %
3	14	18 %
2	10	13 %
1 – Not at all	2	3 %
No answer	2	3 %
Sum	**78**	**100 %**

How should this result be interpreted? When we correlate these findings with the adopted SDM we find no significant associations. It does not affect this profile whether the SDM is the company's own or a commercial method, neither if the company uses a formal SDM nor only a set of techniques.

We interpret this result as an indication that the two disciplines currently seem to live side by side. They are not integrated, neither are they perceived as contradictions. Thus, we find some support also for our second hypothesis.

Do these findings support the somewhat pessimistic view from several usability researchers [3], [7] that the two cultures are irreconcilable? We think the answer is no, for two reasons. First, we have documented that most IT companies do not view formal SDM as rigid frameworks; rather they pick and use elements that integrate with

their existing work practices. This situation makes it much easier to also integrate usability work. The second reason is that the IT companies in this survey do view usability as a key factor for project success. What is lacking is probably a clearer role for usability work, as also suggested by Boivie et al [3].

4.4 Limitations

We acknowledge that there are limitations to this research. The 259 companies do not represent a random sample of the IT company population, which may bias our results. Regarding the questionnaire, one may question whether the respondents have the same understanding of the usability terms as in the IS research community. Further research should address these issues.

5 Conclusions

This paper investigated the adoption of systems development methods and usability, through a web based survey in the software industry in Norway. The significance of this research is that it extends earlier case study research on SDMs and usability, within an industrial perspective. Although our sample is not fully statistically controlled in relation to the population, we argue that it is large enough to justify the findings.

The point of departure was the assumption that two important practices in software development, one of traditional systems development methods and one of usability work, are not integrated in industrial software projects.

We find that the use of systems development methods is fairly stable in Norway, confirming earlier research. Most companies do not use a formal method, and of those who do, the majority uses their own method. Generally, the use of methods is rather pragmatic: The companies that do *not* use SDMs report that they use elements from such methods. Further, companies that use their *own* method import elements from standardised methods into their own.

We find support for our first hypothesis; that companies pay lip service to usability but do not prioritize it in development projects. This applies particularly to usability testing. We also find some support for our second hypothesis; that systems development methods and usability are perceived as not being integrated.

These finding do not, however, support a view of two cultures of systems development. Both the flexible approach to systems development practices and the generally positive attitudes to usability allow for a gradual integration of usability techniques into traditional systems development.

References

1. Beck, K.: Extreme Programming Explained: Embrace Change. Addison-Wesley, Boston (2000)
2. Beyer, H., Holzblatt, K.: Contextual design: A customer-centered approach to systems designs. Morgan Kaufman Publishers, San Francisco (1997)
3. Boivie, I., Gulliksen, J., Gøransson, B.: The lonesome cowboy: A study of the usability designer role in systems development. Interacting with Computers 18, 601–634 (2006)

4. Bygstad, B., Fagerstrøm, A., Østensen, T.: Exploring the relationship between software development processes and IT based business innovation. A quantitative study in Norway. In: Proceedings of NOKOBIT 2004: Norsk konferanse for organisasjoners bruk av IT, Stavanger, Norway (2004)
5. Bygstad, B., Fagerstrøm, A., Østensen, T.: Vi bruker vår egen metode, og den fungerer utmerket. En undersøkelse av hvilke utviklingsmetoder som er i bruk i norske programvareutviklingsmiljøer. In: Proceedings of NOKOBIT 2002: Norsk konferanse for organisasjoners bruk av IT: Kongsberg (November 25-27, 2002)
6. Fitzgerald, B.: An empirical investigation into the adoption of systems development methodologies. Information & Management 34, 317–328 (1998)
7. Iivari, N.: 'Representing the User' in software development - a cultural analysis of usability work in the product development context. Interacting with Computers 18, 635–664 (2006)
8. Jacobson, I., Booch, G., Rumbaugh, R.: The Unified Software Development Process. Reading, Addison Wesley (1999)
9. Larman, C.: Agile and Iterative Development: A Manager's Guide. Addison-Wesley, Reading (2004)
10. Larman, C., Basili, V.R.: Iterative and Incremental Development: A Brief History. IEEE Computer 36(6), 47–56 (2003)
11. Microsoft, Microsoft Solutions Framework (2004)
12. Nielsen, J.: Usability Engineering. Academic Press, Boston, MA (1993)
13. Sommerville, I.: Software Engineering. Harlow, Pearson Education (2001)
14. Stapleton, J.: DSDM: Business Focused Development. Addison-Wesley, Harlow (2003)
15. Truex, D., Baskerville, R., Travis, J.: Amethodical systems development: the deferred meaning of systems development methods. Accounting, Management and Information Systems 10, 53–79 (2000)

Activities for Usability in Lenovo China

Baihong Chen and Rong Yang

User Research Center, Lenovo Corporate Research China, Beijing, P.R. China
chenbh@lenovo.com, yangronga@lenovo.com

Abstract. This paper briefly introduces activities for usability in Lenovo China. First, the progress and the current status of usability in Lenovo China will be introduced. Second, a basic flow of usability activities is summarized. Finally, some results of activities for usability applied to Lenovo products will be presented.

Keywords: usability, design process.

1 Introduction

With the transition from product economy to service economy and then to experience economy, more and more companies begin to strongly sense the importance of transforming the conception of *product-oriented* into *user-centred* in the product design. The academic circles also put up forward some theories and methods about it. From the 80th on, the idea of *user-centred* gradually penetrates through the whole life cycle of a product. The *user-centred* is to concentrate on the users' characteristics and tasks, in order to ensure that product functions meet needs of user, and to ensure that the operation of products comply with the users' habits in the expected context of use(Beyer & Holtzblatt, 1998). Improving the *quality of use* of products has been the ultimate goal of the usability engineering.

Lenovo is the biggest IT Company in China market from 1996 to now, and becomes the third PC Company in the world after the merge of IBM PCD in 2004. New Lenovo pays more and more attentions on user experience in the whole product lifecycle. Therefore, how to improve usability quality is one of the most important effort directions of new Lenovo.

Actually, in later of 2000, Lenovo was one of pioneers planed to develop user experience research and practice in China. In 2001, a technical team named URC (user research center), was set up in Lenovo corporate research to carry out activities for innovation and user experience. After 6 years' development, the team of URC plays some important roles in concept design, HCI research, prototype design, and usability evaluation etc., and many research results have been applied to Lenovo mobile phone, desktop, and laptop. Among these roles, UCD (User-centered Design) and usability are the fundamental activities for a lot of product development in Lenovo China. This paper will introduce some activities for usability and their contributions to Lenovo products, and introduce our thinking for the future application of usability in Lenovo also.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 267–273, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 Frame and Contributions of Usability in Lenovo China

As one of important product qualities, usability quality should be ensured in many design and development phases and should involve user information. User involvement in the product design process has been an important topic in UCD and usability field. Generally, there are two types of user involvement. One is *user-driven design*, which means direct involvement of potential users in the design process. The other is *user-informed design*, which means involvement into the context of use by some methods such as usability test, focus group or field study. The former approach requires considerable user involvement over a long period, which is often not applicable to more *user-oriented* products with their relatively short product development times. Moreover previous experiences showed that the user is not always the good designer (Nielsen, 1993), so the *user-driven design* is few adopted. At present, the *user-informed design* is often applied in the design process, and usability test or evaluation is one of popular methods of obtaining needs or feedbacks of user for product design (Wiklund, 1994; Mandel, 1997, Mayhew, 1999).

2.1 Frame of Usability

After many years' practices in different products such as hardwares or softwares, desktop or laptop, even mobile phone, we have established some basic usability workflows to satisfy different requirements. These flows include target user research, interaction research and design, visual design, test and evaluation, and even user trace after product released. As a sample, figure 1 shows a basic usability workflow for the product lifecycle.

Like many other flows, some other infomation must be inputted into usability workflows. These information can be classified as product and technology information, marketing information, and user requirement information. Technique department collects product and technology information, which includes product idea and concept, technology trend and technique specification and so on. Planning department and social research department gather the market information, which includes company decision and target user segmentation.

In the figure 1, input information includes technology information and marketing information. They are as the input in the process of elicitation and validation of requirements, whereas these two kinds of information would influence with each other. The main usability workflow has four basic phases, including user reserch, interaction and usability research and design, and visual design and feedback analysis. Each phase has its own outputs according to differnt product requirements. While the outputs of user research can be the inputs of interaction research and design, and the outputs of interaction can be the inputs of visual design also. In the figure 1, part A can be used to obtain some clues for innovation, and part B is more detailed for improvement of usability quality.

There may be required several usability tests or evaluations during the product lifecycle. Nielsen(1994) summarized four sorts of usability test, including exploratory test, assessment test, validation test and comparison test during the whole design process. And many researchers and practicer developed a lot of effective usability methods (Nielsen, 1994). Here, we don't introduce usability methods.

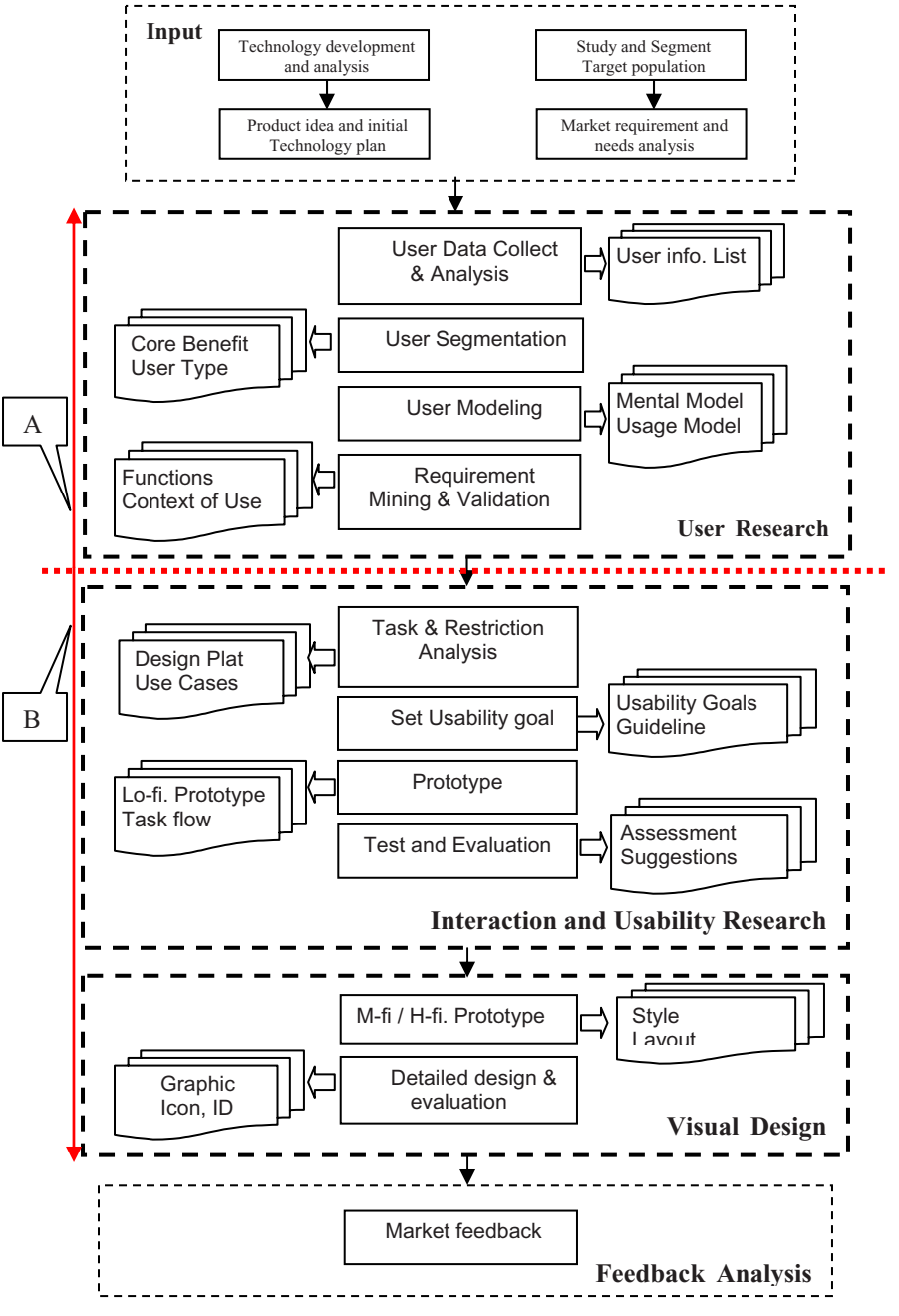


Fig. 1. A basic usability workflow for product development

2.2 Contributions to Business

For a company, it is obviously important to contribute to business. There are two perspectives to consider:

Improvement of Usability Quality. Markets for most products are becoming more competitive. Usability is becoming more important than ever. Users are demanding usable products. At the same time that user expections are rising, developers are being challenged to produce products on shorter schedules at lower cost. In this environment, techniques that are very efficient and effective are needed. Usability research in Lenovo is meeting that challenge, and activities for usability are giving more and more inputs to improve the usability quality of products. In the figure 1, part A and part B are often used as a whole to contribute to usability quality, while part B is used more than part A here.

Elicitation of Innovation. How to meet the needs of user and give them more conveniences to operate a product needs many different research results. While in implement of usability engineering, we can obtain much information on needs of user and context of use. This kind of information can bring designers the elicitation of innovation, and products with this kind of innovations always match the real or even potential needs of user. New Lenovo is a creative company, and activities for usability can elicit innovation of products. In the figure 1, many clues for innovation can be obtained after part A finished.

3 Cases Study

From 2001, activities for usability were carried out on different products such as products of desktop, laptop and mobile phone. Different degree of usability test, evaluation or consultation has been provided for almost 70 projects or teams. The following two cases, one is applied to Lenovo mobile phone to improve the usability quality, the other is applied to Lenovo desktop to elicit innovation, are introduced respectively.

3.1 Case 1: Improve Usability Quality for Mobile Phone

The usability performance of mobile product influences its competition in the market. After practices in different products, interaction specifications for Lenovo mobile phone, which include menu trees, function icons, idle interfaces, logic flows for typical functions, etc. have been established according to different targets users.

These specifications provide effective guarantee for the usability performace of mobile phones. For example, figure 2 shows that the different operation steps of SMS between before and after usability research. After analysis and feedback of usability test of SMS operation, the final operation steps curtailed from 7 to 3 steps. Presently, the operation step of SMS in Lenovo mobile phones is one of the simplest one in mobile phones appeared in China market.



Fig. 2. Operation steps of SMS, the above is the original steps, the below is the optimized steps

Besides the SMS, activities for usability in mobile phone also improve the operation performance of other functions such as camera, music etc. And usability researches for mobile phone spread from the efficiency researches to the emotion researches. These researches also give more and more inputs for Lenovo mobile phones and improve their competition in China market.

3.2 Case 2: Elicite Innovation for Desktop

Lenovo desktop has won the biggest market share in China since 1996. There are two main reasons: one is deep understanding of China desktop users and many series of desktop products developed with some advanced and practical functions, the other is knight sarvice.

Usability research is one of the most important parts of the former reason and ensures the ease of use for desktop products. For example, one key for internet access was a practical appliaction and provided an easy way to internet access for ordinary internet users in 1999.

And usability research also enlarges the innovational space for desktop. Figure 3 shows an example of innovation via usability research. In 2004, Lenovo released a game desktop named Fengxing with three convenient functions for game user. These functions provide mass data exchange among game users via detachable hard disk,

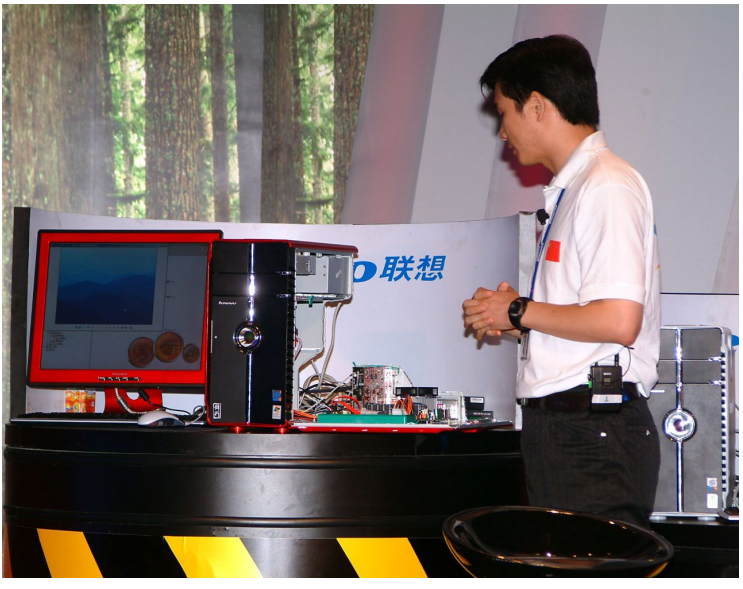


Fig. 3. Desktop with detachable hard disk, easy cover, and mode-switch

and convenience to update new components such as game card or memory via easy cover, and speed-adjust of desktop via mode-switch.

4 Discussion

We are now enforcing UCD and usability in our company. Our developing people are becoming more usability-minded year by year, and usability technologies and tools are to be updated as the usability level of developers goes up. However, we must make the change as discrete as possible and define the usability level scale accordingly. We need more natural usability guidelines for diverse products developers that reflect the nature of products such as desktop, laptop and mobile phone, or software or hardware. And we must make our UCD and usability process more essential and natural process.

5 Summary

This paper briefly introduces activities for usability in Lenovo China. Since 2001, usability technologies have been applied to many different Lenovo products such as desktop, laptop and Mobile phone. Some UCD and usablity workflows are also established in Lenovo China and these flows can give certain input and output for development teams. The results show that these activities can improve or ensure the usability performance of products remarkably. In addition, some innovational ideas can be obtained via usability research also.

References

1. Mayhew, D J.: The usability engineering lifecycle. Morgen Kaufmann Publisher Inc., San Francisco (1999)
2. Mandel, T.: The elements of user interface design. John Wiley & Sons Inc, New York (1997)
3. Nielsen, J.: Usability engineering. Morgen Kaufmann Publisher Inc., San Francisco (1993)
4. Rubin, J.: Handbook of usability testing. John Wiley & Sons Inc, New York (1994)
5. Wiklund, M E: Usability in practice. Academic Press Inc., San Diego (1994)
6. Beyer, H., Holtzblatt, K.: Contextual design: Defining customer-centered systems. Morgan Kaufmann, San Francisco (1998)
7. Nielsen, J., Robert, L M.: Usability Inspection Methods. John Wiley & Sons, New York, NY (1994)

The Cultural Usability (CULTUSAB) Project: Studies of Cultural Models in Psychological Usability Evaluation Methods

Torkil Clemmensen[1] and Tom Plocher[2]

[1] Department of Informatics, Copenhagen Business School, Denmark
tc.inf@cbs.dk
[2] Honeywell Labs, Minneapolis, USA
tom.plocher@honeywell.com

Abstract. Cultural models in terms of the characteristics and content of folk theories and folk psychology have been important to social scientists for centuries. We suggest that they should be at the heart of the scientific study of human-computer interaction (HCI). The CULTUSAB project is conducting an in-depth investigation of the key dimensions of culture that affect usability testing situations, including language, power distance, and cognitive style. All phases of the usability test are being evaluated for cultural impact, including planning, conducting, and reporting results. Special attention is being focused on subject-evaluator communication and cultural bias in the test design and structure of the user interface being tested. Experiments are being replicated in three countries: Denmark, India and China. The research will result in new testing methods and guidelines that increase the validity of usability tests by avoiding cultural bias, and allow us to produce comparable results across different countries.

Keywords. Cultural usability, think aloud usability test, cross cultural research.

1 Introduction

The CULTUSAB project aims to investigate the impact of culture on the results of established methods of usability testing. Cultural models in terms of the characteristics and content of folk theories and folk psychology have been important to social scientists for centuries. From Wilhelm Wundt's Volker psychology to the distributed and situated cognition theorists in the global world of today, thinkers have seen human action as being controlled by cultural models. Cultural models for humans interacting with computers should therefore be at the heart of the scientific study of human-computer interaction (HCI).

Historical imperatives aside, there are numerous indications from practical experience that usability testing procedures developed for use in Europe or the US do not necessarily give the same results when applied in India or China. The CULTUSAB project is conducting an in-depth investigation of the key dimensions of culture that affect usability testing situations, including language, power distance, and cognitive style. All phases of the usability test are being evaluated for cultural

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 274–280, 2007.
© Springer-Verlag Berlin Heidelberg 2007

impact, including planning, conducting, and reporting results. Special attention is being focused on subject-evaluator communication and cultural bias in the test design and structure of the user interface being tested. Experiments are being replicated in three countries: Denmark, India and China. The research will result in new testing methods and guidelines that increase the validity of usability tests by being sensitive towards cultural bias, and allow us to produce comparable results across different countries. This 3-year research project started in May, 2006, sponsored by a grant from the Danish Research Council to Copenhagen Business School. India Institute of Technology-Guwahati, the Chinese Academy of Sciences, University of Copenhagen and Roskilde University are co-investigators. Industry advisors to the project are from Honeywell, Nokia, Human Factors International, and Snitker Associates.

2 Motivation and Purpose

With the advent of globalisation and the information technology revolution in developing countries, we can no longer overlook the aspect of culture in the design of user interfaces and interactive products. We need to understand and accept that there are significant differences in how people with different cultural backgrounds respond to directions and test methodologies. From the lab of the very large IT companies in Beijing to the design departments at India's finest institutions of higher education, there is a call for adequate methods and techniques for designing human-computer interaction. Usability issues such as how to support input to computers of the many Chinese characters within a classic windows-and-mouse paradigm, or how in a multilingual and multicultural country as India to design and evaluate the usability of interfaces to Automatic Teller Machines and other text-based interactive products must be addressed. In Denmark and Europe we face even more complex challenges to the quality of information and communication technology, as our societies turn increasingly multicultural and we must provide networked information to both ethnic majorities and minorities. Despite of these challenges, we do not have any kind of formal methods which guides us to evaluate a product to a certain standard while keeping sensitivity to cultural issues. With this project we take as our point of departure the issues of how to avoid cultural bias in requirements elicitation and usability data collection. Which user-based evaluation methods do address cultural diversity in both the moderator and user? We study the fundamental and widespread assumption that the usability evaluator needs to have the same cultural background as the test user in order to completely understand how users will respond to the test instructions and test methodology. Furthermore, we aim to understand what the effective way to obtain test users' usability feedback is, without actually disguising the usability problems. Our research question is:

What is the impact of culture on the results of established methods of usability testing?

- How are the different components of a usability test, e.g. planning, performing and reporting, influenced by a cultural diversity of users and contexts of use?
- How are cultural backgrounds taken into account when recruiting and describing usability test users?

- Which form of relations and communications between evaluators and test users are most effective in terms of finding relevant usability problems in culturally localized applications?
- What is the nature of common cross culturally related usability problems, and what is a good quality of cultural usability of information and communication technology?

2.1 Cultural Usability and Usability Evaluation Methods

A focus in this project is Usability Evaluation Methods (UEMs), as defined by [7][1]. In the industry, a wealth of UEMs is used to evaluate computer software user interfaces and other interactive products: Inspection methods, Workplace observation, Think-Aloud Usability Test, etc. Both in the industry and in research there is an interest in understanding cultural issues because there are many cultural factors that influence usability evaluation results. Some have to do with culturally biased guidelines and procedures in using a specific UEM, while others are related to other types of cultural differences appearing in test situations. There is an entire spectrum of factors ranging from those completely independent of the UEM to those that are practically built-in in a particular UEM. For reasons of comparability, the project needs to consider more than one type of UEM.

The theoretical part of the project will analyze the concept of 'cultural usability' through analysis of the use of UEMs within a cultural and social diversity of users and contexts. The international diversity of users and contexts of use is an expansion of the traditional usability research, which is based on more simple, regionally specific conceptions of users [4, 11]. The research methodology, mostly qualitative, allows for in depth investigation of the conceptual and practical layers of user and context representations in established UEMs.

3 Background

The discussion about culturally localised interfaces has been fairly conclusive on the point that localization is not just mere translation of text, it's more than that [14, 15]. To locally adapt user interfaces, we must use usability engineering methods similar to those used in the development of original user interface. However, the existing practice derived from the West of migrating software from a source culture to a target culture may work in the design and implementation phase, but not in the usability evaluation phase [22]. For example, in Malaysia having a test user of higher rank than the experimenter will result in more negative comments about the product than having a test user of lower rank than the evaluator. In some countries testing subjects individually should be avoided, as little information may be retrieved [8]. In an interview study done in India, those participants with a similar socio-cultural background as the interviewer (India) brought more usability problems than participants who were interviewed by the interviewer with a different socio-cultural

[1] UEMs is a broad term for analytical and empirical methods that usability professionals use to evaluate the interaction of the human with the computer with the purpose of identifying aspects of this interaction that need to be improved to increase the usability of the product.

background (in this case Anglo-American) [19]. Others have raised similar issues: Do language and cultural differences between staff and participants negate the outcome of usability tests? Are foreign nationals good representatives of users in their home country? These practical issues are of great importance to the design and use of usability evaluation methods.

The background for many studies of cultural aspects of usability are Hofstede's cultural dimensions [9]: power distance, individualism-collectivism, masculinity-feminism and uncertainty avoidance. Most culture and design theorists, many professional designers across all disciplines and also some users believe that these cultural dimensions pervade every human activity and every artifact, including user interfaces [13]. Recently, however, opponents of this approach argue that the current process for the design of universally usable systems is not appropriate, because of its overdependence on guidelines, difficulty of determining the user from the present cultural grounds, its tendency to build stereotypes which later become design rules, and its treatment of different cultures with one specific language that doesn't take into account cultural heterogeneity.

Instead, these researchers see culturally determined usability problems in interfaces as caused by the users' (mis-) understandings of the representations whose meaning lie in the culture-specific context [1]. This conceptualization of cultural usability is in line with more recent social psychological approaches to culture that take into account the establishment of 'social facts' and peoples' sense of the 'reality' of social groups, and see these as effects of peoples' use of symbols to construct their social reality; processes that again are firmly related to culture and communication [12]. These processes are important for cultural usability. For example, in our pilot studies in India and Denmark of the thinking aloud usability test method [5, 21], the test users quickly realized that some test evaluators did not belong to the user's own social group, and acted accordingly by explaining to the foreign evaluator aspects of the test application that would seem to be obvious and not require explanation to an evaluator from the same group. In the end, this meant that some relevant usability problems were not identified due to cross cultural issues.

4 Approach and Method

We base our approach on a moderate universalism [16]: 1) maybe there is cross cultural universal usability, maybe there is not, we need empirical documentation, 2) universal usability will most probably be found on the level of theoretical principles rather that phenomena, and 3) we need to make assumptions about universal usability to help organize data into general theories. With this we look away from the two sisters of universalism [17]: evolutionism (one society is more advanced than others) and relativism (societies must be understood from their own perspective) approaches, in order to create the best ground for comparability of results and collaboration among the researchers in the project.

4.1 Social Psychological Approach to Cultural Usability

In the study of UEM in a cross cultural perspective, we suggest to apply a social-cognitive model of culture [10] that conceptualize culture as a loose network of

domain-specific cognitive structures (including theories, beliefs), and, furthermore, argues that an individual can hold more than one cultural meaning system, even if the systems contain conflicting cultural theories. Depending on the accessibility, availability and applicability of such cultural knowledge, cross-cultural differences may impact usability.

Accessible cultural knowledge can be approached as meaning systems that are widely shared among members of a cultural group and frequently used in communication among members and thus becomes chronically accessible. In a usability test situation, where people under time pressure look for readily available and widely accepted solutions to a problem, the chronically accessible knowledge will be used and typical cultural group differences will emerge.

It is however not sufficient to have task conditions that favor the use of chronically accessible cultural knowledge. Since individuals in a society increasingly are poly-cultural in their background and thus have more than one implicit theory of how to perceive and act in a given situation, the individual chooses or implicitly applies the theory that is most accessible in that situation. Therefore, in the study of UEMs it can sometimes be necessary to ensure the availability of culturally accessible knowledge by including ways to activate or 'prime' this knowledge. Such primers can be cultural icons and pictures. For example, we can test localized IT applications that contain culturally specific icons and pictures that can prime evaluators' and test users' culturally specific knowledge systems, while they complete a behavioral strategy such as a think aloud usability test.

In our approach, we suggest to deal with the assumption about appropriateness of applying cultural knowledge by pairing evaluator and users of different respectively similar socio-cultural backgrounds. In order not to miss significant parts of the social realities of a postmodern world [2], we can study UEMs that are performed at different 'home grounds' such as China, India and Denmark. A great variability in sub-studies will be needed in order to estimate the universality of claims about cultural usability in the project. The glue that can bind such sub-studies together will have to be that individual researchers are present at the studies and field experiments which are done at the other researchers' home grounds.

5 Expected Outcomes of Cultural Usability Research

5.1 Practical Application of Results

Studying cultural usability will have significant societal impact on issues related to cultural aspects of interaction design and usability testing. Local usability professionals will improve their understanding of usability in other parts of the world and their ability to configure usability evaluation methods cross culturally in other nations or in ethnic minority settings within the nation or region. An understanding of the cultural aspects of usability will help the designer and developers to analyze the ontology of the application domain of a system by revealing the semantics of the domain from the users' many points of view. The openness of the technology for a wide range of interpretations makes it very important to develop UEMs that help the designers and developers to investigate the use of technology on many levels of detail

within society. This is very important in current efforts in coordinating between incompatible system developments methods such as the natural science-oriented object-oriented analysis and design approach versus the humanities-oriented interaction design approach to usability [6].

5.2 Publication of Results

Results of cultural usability research should be publishable in high level international HCI journals such as Interacting with Computers, Behaviour & Information technology and International journal of Human Computer Studies. One obstacle for publication may however be the need for a cross-cultural research design which makes the research more complicated to communicate. Before being published in journals, findings may have to be presented and discussed with researchers and industry at appropriate national conferences such as the annual Danish HCI research symposium, the INDIA HCI conferences and HCI International 2007, Beijing.

5.3 International Collaboration and Methods Development

In developing the methods of testing intercultural usability evaluation it is at the same time necessary to develop and evaluate the methods for doing so. Moreover, the intense collaboration between HCI researchers from different regions of the world in specific projects, from field testing to analysis and publication, will strengthen research networks between the countries involved and pave the way for future research in this and related areas which will benefit all the participating research institutions and researchers and their students, see e.g. [18] for an example. Opportunities to cooperate about research in cultural usability with HCI researchers from the emerging HCI communities in developing countries across the world should be exploited rigorously, e.g. [3, 20].

5.4 Educational Benefits

The educational significance of the project lies in that students will benefit from the global perspectives on human-computer interaction. While of potential interest to all HCI students of today, Cultural Usability research should be of special interest to students having a multicultural background or expected career in a multicultural environment.

5.5 Future Research Agenda

The next research in cultural usability may focus on the training of users as part of improving the usability of information and communication technology. As we know from numerous studies, there are high costs associated with learning to use new systems and with the social psychology of the surrounding cultural and communicative processes.

Acknowledgements. This paper was written partly on a grant from the Danish Council for Independent Research to the Cultural Usability project.

References

1. Bourges-Waldegg, P., Scrivener, S.: Meaning, the central issue in cross-cultural HCI design. Interacting with Computers, 9(3), 287–309 (1998)
2. Burawoy, M.: Global Etnography. California Press (2000)
3. Cecilia, M., Baranauskas, C.: HCI in Brazil: Prospect and Challenges. in INTERACT 2003 - Bringing the Bits together - Ninth IFIP TC13 International Conference on Human - Computer Interaction, Zürich, Switzerland, pp. 1081–1083 (September 1-5, 2003)
4. Clemmensen, T.: Four approaches to user modelling - a qualitative research interview study of HCI professionals' practice. Interacting With Computers, vol. 16(4), pp. 799–829
5. Clemmensen, T., Goyal, S.: Cross cultural usability testing. Working paper, Copenhagen Business School, Department of Informatics, HCI research group, 2005-006. 20.
6. Clemmensen, T., Nørbjerg, J.: Separation in Theory - Coordination in Practice. Software Process Improvement and Practice, vol. 8, pp. 99–110
7. Gray, W.D., Salzman, M.C.: Damaged merchandise? A review of experiments that compare usability evaluation methods. Human-Computer Interaction, vol. 13(3), pp. 203–261
8. Herman, L.: Towards Effective Usability Evaluation in Asia: Cross-cultural differences. In: OZCHI 1996 (1996)
9. Hofstede, G.: Geert Hofstede's Homepage (2005)
10. Hong, Y.-y., Mallorie, L.M.: A dynamic constructivist approach to culture: Lessons learned from personality psychology. Journal of Research in Personality, vol. 38, pp. 59–67
11. Isomaki, H., Pohjola, A.: Introducing Multiple Views on Gender and Information Technology. In: Isomaki, H., Pohjola, A. (eds.) Lost and found in virtual reality: Women and Information technology, University of Lapland (2005)
12. Kashima, Y.: Culture, Communication and Entitativity - A Social Psychological Investigation of Social Reality. In: Yzerbyt, V., Judd, C.M., Corneille, O. (eds.) The psychology of group perception - perceived variability, entitativity and essentialism, Psychology Press, NY (2004)
13. Marcus, A.: Fast forward: Culture class vs. culture clash. Interactions, vol. 9(3), pp. 25–28
14. Molich, R., Dray, S., Siegel, D.: Tips and Tricks for a Better International Usability Test. in CHI, Special Interest Group (Vienne Austria, April 24-29, 2004), ACM (2004)
15. Nielsen, J.: Designing for International Use. In: CHI 1990 (1990)
16. Pepitone, A.: A social psychology perspective on the study of culture: An eye on the road to interdisciplinarianism. Cross-Cultural Research, vol. 34(3), pp. 233–249
17. Shweder, R.A., Bourne, E.J.: Does the Concept of the Person Vary Cross-Culturally? In: Shweder, R.A. (ed.) Thinking Through Culture - Expeditions in cultural psychology, Harvard University Press, London (1991)
18. Smith, A., Prasad, S.: IESUP - Indo European Systems Usability Partnership, Bangalore, India, Project website (2004)
19. Vatrapu, R.: Culture and International Usability Testing: The effects of Culture in Structured Interviews. Master thesis. Virginia Polytechnic Institute and State University (2001)
20. Wesson, J., Greunen, D.v.: New Horizons for HCI in South Africa. In: INTERACT 2003 - Bringing the Bits together - Ninth IFIP TC13 International Conference on Human - Computer Interaction, (Zürich, Switzerland, September 1-5, 2003), pp. 1091–1095 (2003)
21. Yeo, A.: Global-software Development Lifecycle: An Exploratory Study. In: CHI 2001 (2001)

Cultural Usability Tests – How Usability Tests Are Not the Same All over the World

Torkil Clemmensen[1], Qingxin Shi[1], Jyoti Kumar[2], Huiyang Li[3], Xianghong Sun[3], and Pradeep Yammiyavar[2]

[1] Department of Informatics, Copenhagen Business School, Denmark
{tc.inf,qs.inf}@cbs.dk
[2] Indian Institute of Technology Guwahati, Assam, India
{jyoti.k,pradeep}@iitg.ernet.in
[3] Inst. of psychology, Chinese Academy of Science, Beijing, China
{lihy,sunxh}@psych.ac.cn

Abstract. The cultural diversity of users of technology challenges our methods for usability evaluation. In this paper we report on a multi-site, cross-cultural grounded theory field study of think aloud testing in seven companies in three countries (Denmark, China and India). The theoretical model that emerges from the data suggests that the production of a usability problem list is multi-causal and subject to cultural variations. Even the way usability problems are experienced by test participants may be different. In the discussion we outline practical guidelines for a test that is more sensitive towards cultural usability.

Keywords: Usability test, think aloud, cultural usability, field study.

1 Introduction

Culture plays an increasing role in discussions of information and communication technology. As of today, we do not have any formal methods to guide us in evaluating a product to a certain standard while being sensitive to cultural issues. Cultural usability tests are not yet established methods. Thus, in this paper our point of departure is a look at the methods that we have and to consider the following research questions. Which current practices of think aloud (TA) usability testing address cultural diversity for both the evaluator and the user? How do test users respond to the test instructions and test methodology? Which interfaces are more influenced by cultural diversity and current usability test practices? We try to understand the most effective way to get usability feedback from test users' without actually disguising the usability problems.

The rest of this paper is organized as follows. The following section deals with the multi-site cross-cultural method, section three presents the theoretical model that emerged from the observations, and section four discusses our findings.

2 Method

We approached the research questions by studying current TA usability testing practice in companies that test software and interactive products for the local market

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 281–290, 2007.
© Springer-Verlag Berlin Heidelberg 2007

in three countries (China, Denmark, and India). The selection of participants was done on the basis of global ethnography [3] which suggests that we need multiple studies in different settings to shed light on a phenomena. We used a site-based procedure [1] for locating and recruiting qualitative study participants to learn about each cultural setting and to gain entry into the setting. Our procedure had five steps:

1. Specification of the characteristics relevant to the sample: geographic (Denmark, India, China), socio-cultural (experienced moderators/evaluators and local test users taking part in TA tests), company characteristics (we only used companies that did professional user testing services of software products to the local market)
2. Generation of a list of sites - the places where TA tests were done
3. Estimate of the composition of clientele at each site by contacting a 'gatekeeper' for each site and asking for appropriate statistics for the site and helping to gain entrée; the 'gatekeeper(s)' in our case were the managers in each company who had the daily responsibility for running TA tests'
4. Recruitment of participants and 'gatekeeper' (in reality, the manager) and an agreement on when and how the TA tests at the site could be observed and the test users and moderators could be interviewed
5. Recruitment of individuals from sites and maintenance of a table indicating the characteristics of the participants in the sample [13] to help the researcher to assess the quality of recruitment. In our case it revealed that we had to live with demographic between-site differences between test users and moderators

We achieved a sampling diversity that was saturated in the sense of: a) we were not able to get into contact with more willing TA test vendors; b) we did achieve a reasonable amount of variation in our sample; and 3) it was quite clear that the three geographical categories of TA test vendors were clearly independent (none of the companies cooperated). In each company, we did field observation with video cameras of TA usability test sessions and afterwards interviewed the evaluators and the test users. We were three observers/interviewers: an Indian, a Danish and a Chinese. We made sure that two or more were present at all observations to increase cultural validity. We made it explicit that the TA test was to be run appropriately; i.e., the usability test manager should make the decisions about how to run the usability test to allow us to observe the current practice of TA in that company. The test application should be aimed at the local market. If it was not possible to observe a customer-paid TA test, we asked the company to redo a recently run customer-paid test. This procedure gave us a total of 52 hours of observation in three languages across seven companies. The analysis of the interviews was done through a grounded analysis approach [7] in which we focused on the production of usability problems in the conduct of think aloud usability tests.

3 Results

The grounded theory model for conducting a think aloud usability test was based on previous conceptual and empirical work within the cultural usability project [4], and then developed further from the field studies presented in this paper, as shown in Figure 1. The goal of the analysis was not to have an accurate description of all data, but a quest for a conceptual theory abstract of time, place and people [7].

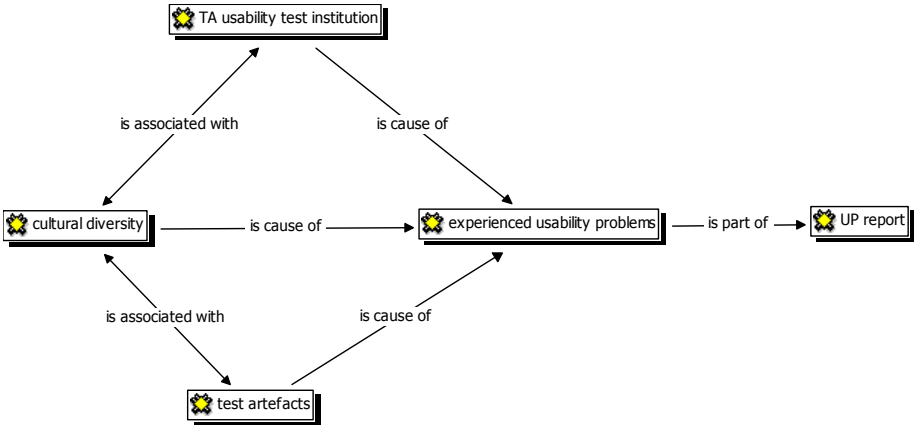


Fig. 1. Theoretical model for conducting think aloud usability tests

3.1 Cultural Diversity in Moderators and Test Users

The two traditional independent variables of: 1) age and 2) gender emerged from the field data as important parts of the construct of cultural diversity among moderators and test users. The Indian moderators saw age and gender differences between moderator and test users as no problem with urban users, albeit as a potential problem with rural users of a traditional cultural background where it was important to "speak their language", "be polite and respectful towards the elderly", and "be keen to listen". Specifically, a senior moderator could frighten a young test user with a traditional cultural background and require the moderator to "go an extra mile to communicate". Even in a usability test conducted remotely in which the moderator would not see the test user, the moderators would want to know the age of their test users. Additionally, if the user were a female from a traditional family, a male member of her family had to be present during the pre-interviews and the test, or the moderator needed to be a woman, (moderator): "It takes a woman to interview female users with traditional background (female members of a traditional family)". In the Chinese tests, age and gender differences were emphasized, (moderator): "….the way they do the test will be different…". "…if the user is male it is better if the moderator is a female…". In the Danish tests, the participants were positive towards the test users who were most similar to them in terms of age, gender and job experience (audience): "He was a very good test user", and hostile to the extent of being offensive towards a test user who was different in age and gender (audience): "She is an old …..."; at one point the moderator could hear the audience laughing and he was afraid that the user should (moderator): '…play this…'.

3.2 The Usability Test as an Institution

The cultural diversity of test participants explains partly the usability test practice observed in the three countries in terms of variations and additional properties of established usability tests [2]: 1) specific test goals and concerns, 2) real work tasks to be done, 3) think aloud procedures, 4) test users, evaluators, observers, clients/managers,

designers, in-house trainees, 5) recovery procedures, 6) evaluator room and observer room, 7) test applications and usability problem lists/reports. All incidents of usability tests that we observed did share this general approach to usability testing, and most exhibited interesting variations of the properties of the test.

Specific test goals and concerns. In the Danish tests, one goal was to get the test users to think aloud to allow insights into the test user's cognitive work with the test application. However, from the beginning it was unclear exactly when the user was required to think aloud. The evaluator asked about the user's opinion and preferences most of the first part of the two hours that each test took, and then gave the user a more specific task to do towards the end that was more like think aloud. In the Chinese tests, it was also a goal to do a think aloud test. All the moderators said it was a think aloud test, although some (moderator B + S) pointed out that in their case the test was actually a combined think aloud and interview. Moderators asked questions and the users answered throughout the whole session in all the studied companies. Furthermore, upon occasion the users did the tasks silently, and then after the task completion the moderator probed or asked for explanation, i.e., no think aloud occurred. The reason for this non-interference from moderators was given as (moderator M): "… there is no need to disturb the user when she takes the correct steps….". The user did not seem to think aloud actively in the Chinese test. In the Indian tests, it was an explicit goal with the scenario that the test user should be open and tell about his or her mind. However, all moderators were obviously concerned that usability testing should be seen as part of a user centered design process, (Moderator 3): "Basically the final outcome for us has to be a tangible physical design. So for that we get more detailed information…".

Real work tasks to be done. In the Indian tests, all moderators used a detailed protocol to get the users started on different tasks, (Moderator): "Through scenarios we start the task, when the user has stopped talking and he has answered most to the thing and if I have no questions to ask, the task is stopped", (Moderator): "I just tell them this is the whole scenario, this is what we have to do…to stop: The person has finished all tasks…", (Moderator): "I say: this is what to do, how will you do it…To end: it is a natural end, when the final step has been reached…". It was a concern that the users understood their tasks, which they did (user 1): "I think it is a good task", (user 3): "It was a good task, it made me realize, I found it quite interesting, because I go on sites all the time- I was able to compare with other sites….makes for a healthy ..comparison". In the Chinese tests, task management was informal, both regarding the start and end of the user's task, (moderator): "…lets see the next task…", or simply "After I introduce a task, then it will begin". The concern of having a detailed protocol to guide the test was not always strong, (moderator B) "…if you have many protocols it will scare the user, make the user very nervous, uncomfortable", so that the argument would be that if a moderator needs a protocol, one should ask her or him (moderator B) "Did she do the real usability test or not? In the real usability test, they should not have….it would scare the user...". Another moderator was quite concerned with having and following a protocol, (moderator S): "…sometimes I read out the task in order to make all users get the same instructions…". In the Danish tests, there was a countless number of steps in the protocol, which was partly due to the client wanting a large part of the test application (a website) tested in one long test.

Think aloud procedures. In the India tests, users were actively thinking aloud and speaking out, (User 1 repeatedly) "Now I will do...", and also sometimes thinking in phrases: "I think in today's age…". When necessary, the moderator used reminders to help the user to think aloud. Reminders could have many forms, see Table 1.

Table 1. Reminders to think aloud, from usability tests in Indian company

– please think aloud	– can you say what you are finding
– what are you looking for in this page	– talk to me what are you looking at
– what is happening now	– keep talking
– whatever you like or dislike or you think you can say	– could you tell me more information about what are you doing, you are expecting…

All moderators used hands and arms to make lively gestures to support their speech. The moderators were trained in body language, (manager): "… moderators are trained about the body language…it is a matter of practice that some follow it and some are in the learning phase. The idea is to keep the user comfortable at all times – by communication, body language, other settings etc. …". From the India tests emerged several ways of probing for more information: 1) when the moderator reacts to the users' initiative, e.g., (moderator) "You just said right", (user) " Yes yes I mean this thing…"; 2) when the moderator gives direction to the user as when the moderator asks, "Did you notice that?" and shows something on the screen, and the user says: "No that was my mistake I didn't", 3) when the moderator actively wants to help the users: "Usually in the first task I always ask such questions in order to let the user know what is his job, later I don't do that because then the user would know what he should do", and 4) when the moderator actively wants to dig deeper into the users think aloud, e.g., when a user says "I am not happy with the description…", and the moderator asks: "Why is that?". The difference between asking questions, reminders to think aloud, and probing is not necessarily clear in usability test practice.

In the Chinese tests, reminders to think aloud were used, (moderator UC): "You have finished the task, what did you think just now?" This kind of retrospective reminders appeared to be necessary. In company (UC) the task was a very simple search task and the user was asked to find articles in the website. It was very easy for her, so she completed the task very quickly (between 2 and 5 seconds per task) and she did not say anything during task performance. In company B, when the user tried to edit (input some words) and was silent for maybe more than one minute, the moderator simply wrote something down on the paper and never reminded the user to think aloud. After a long time, the user said, "I have never done it before….". In company (M), the user did not think aloud at the beginning of the first task (see who called you), and then the moderator reminded the user two times: "…you can talk when you are doing…Which input method are you using now?", and only then did the user begin to think aloud: "…oh I find it, I should fix it, and click the button….". Paradoxically, while almost all users did not consider if the moderator understood them or not, many times after having been silent for some period, they explained to the moderator. The think aloud was an explanation, i.e., the think aloud was retrospective think aloud [5].

Participants. In the Danish tests, the audience consisted of two designers, two marketing people and two managers from the client company. They were obviously important: the senior usability specialist in the company (the head of the usability company) was observing the whole test and developing the usability report in the form of one powerpoint show for each user concurrently with the TA test in front of the audience. Part of the reason for this was to maintain client relations. In the Indian tests, client relations were also important, but the focus was on the relation between the moderator and user, (manager): "The relationship between the moderator and test user should be that of a teacher and a student, the moderator should be the student and the user the teacher (master and apprentice, actually)". The user was supposed to take the role of a design critique who, under guidance of the moderator, evaluates the website. In the Chinese tests, the focus was on the test user-computer dialogue: In two companies, the user did not at all seem to consider the moderator, did almost never look at the moderator, not even when the user said something, e.g., when answering questions from the moderator, the user would still look at the computer screen. Almost all users focused on the task, thinking that the moderator was just a person who gave them instruction and facilitated the test, and they considered him "… just a little more than if I was answering a questionnaire from anonymous sender…". Reasons for this were given during interviews with users to be a) many people in Beijing are very familiar with being interviewed in many situations, b) all users are well educated, know the purpose of the interview, know they are not the subject, c) the users are explicitly told that the test is about the product, not themselves.

Recovery Procedures. A concern in all companies was that the user should not be stuck unnecessarily in a task. For example in the India tests, the moderator helped if necessary, (User 4): "The two times I got stuck he helped me…".

Evaluator Room and Observer Room. In the Danish tests, an observer room accommodated up to 10 clients, observers, trainees, researchers, who through a one-way screen could watch the evaluator/moderator and the test user in the evaluator room. A video of the test user, a PC screen capture and sound were played on monitors in the observer room. A note-taker in the observer room took notes, while the moderator in the evaluator room did not take notes during the session. In the Indian and Chinese tests, there were arrangements very similar to the Danish with a separate note taker in the observer room. In China, Company B was different from the other companies in the sense that there was no dedicated note-taker, the moderator was simultaneously note-taker and wrote the report afterwards. Uncharacteristically, in Company NN the moderator was in the observer room, looking through the one-way screen and interacting through a button-operated telecom with the test user who was alone in the evaluator room.

The Usability Problem Lists/Reports. In the Danish company, the note-taker had the main responsibility for the final usability report to the client. During the tests, he was writing the draft report directly into MS power point, compiling the different users' responses, and structuring the report according to the order in which the test application areas (web site page) were being tested. The moderator was consulted before the note-taker finalized the report to the external client. In the Chinese company B and NN the moderators used a template to write the test report and presented it to the in-house clients. In the case of external clients, a full report was

made. In the Indian company, usually the note-taker used a predefined MS excel template with the test application areas as stated in the test protocol listed in a row and the users in a column, ready to enter the users' reactions, performance scores and satisfaction measures directly into the spreadsheet and then later produce the report. The report usually consisted of 50-80 pages document and a power point slide show.

3.3 The Test Application

What emerged from the observations was that the test users' work with the test application was efficient when they had already been primed with the typical functions of the test application, but were less efficient when they had not been primed. This supports findings from basic research in psychology and anthropology that subjects become fixed on the design function of the object after being exposed to a demonstration of the object's function [6]. It supports the findings from [12] that novice computer users from different cultural groups are not necessarily comparable, but can be seen as relative novices compared to expert computer users with similar cultural backgrounds. Being a novice user in a culture which surrounds you with computers is not comparable to being a novice user in a culture with few computers.

We observed that test users in usability tests will often be urban, modern, young, with higher education, be fluent in English and with substantial computer experience. In the Indian tests, the artifact was a newspapers website and the test users were appropriate users of online news sites: urban, young (20-30 years) with higher education, fluent in English and with experience in using this type of website. In the Danish tests, the test application was a website for an internet and telecommunications provider. The users had higher education, were 25-55 years of age, and end-user of medium to professional expert users of the test application. They spoke their local language, but were also proficient in English. In the Chinese tests, the test application in company UC was an e-learning public school website and the test user a female user, young, not fluent in English, but had experience with similar applications. In company B the test application was a search engine website, and the user was a male user with higher education, not fluent in English, had considerable experience in the test application but not in the tested new functions. In company S the test application was Web-based chat software; in company M Mobile phone interfaces (pen vs key) and in both cases the users were young, female, highly educated, fluent in English and had extensive knowledge of similar applications as test application. In company NN the test applications were: a Web-based work flow tool, with a male user in his thirties, had higher education, was an in-house employee with experience in similar systems, was English speaking, and had a Mobile phone service provider's customer website with a test user who was young and had experience with similar applications. Obviously, in current practice of TA usability tests there is awareness of recruiting users that culturally meet the test applications affordances. One important observation, however, is that when new software is tested, users are not always sufficiently fixed on the design function of the software, i.e., even if the user knows the kind of application being tested, he or she may have little clue about the intended use of the specific functions being tested – while other users may have a clear idea about this. This variation may be one cause of the variations in experienced usability problems.

Finally, some test applications are made for other user groups and require users with other backgrounds like elderly/children and/or rural, traditional, strongly religious, low education, local language only and no computer experience. Furthermore, they may have no or very few computers in their daily environment and hence little general knowledge about what computers can do.

3.4 Experienced Usability Problems and Usability Report

The cultural diversity in moderators and test users, the usability test practice and the affordances of the test application together contribute to the experienced usability problems and to the final usability report to the client. This multi-causal model of experienced usability problems partly contradicts existing usability problem theory that says that the list of usability problems only reflects the properties of the product being tested [11]. However, our study supports the findings that cultural diversity in the test users and moderators [10] and variations in the test setup [9] influence the detection of usability problems. In the Indian and Chinese tests, usability problems were experienced as interactions between test users and moderators, i.e., as co-discovery episodes. In the Indian tests, users when asked, could come up with one or two major problems they believed to have found during the tests, for example (User 2): "…he [the moderator] did give me indications….like where would you go to search for flights….". In the Chinese tests, the user also suggested design changes. For example a moderator asked the user "…why did you not find it right now…", then the user said "…oh I didn't notice this part…", then the moderator said: "…that means they should not put this in this part?" then the user said "…yes they should put it in flash or something…". In Danish tests, the users were quite confident that they could identify usability problems by themselves; the users did not refer to interactions with the test moderator, but instead focused on their own needs as users. These differences could mean the moderator-test user relation extends beyond the session and into the post-session period of fixing the usability problem list.

4 Discussion

Compared to previous studies of cultural usability and usability testing in natural settings, this study is distinctive in its multi–site, cross-cultural approach. A theoretical model of the production of usability problems in seven usability test vendors across three countries was constructed through a combination of interaction analysis observation and grounded theory analysis, which included systematic use of observers with different cultural backgrounds, and checking concepts with the participants to increase validity. This model encompassed cultural diversity in test users and moderators, variations in the conduct of usability test, and assessment of user-technology-fit in a conceptual framework for appreciating nuances in the outcome of usability tests in different regions of the world.

The cultural diversity in the background of users (and moderators) suggests that all usability is culturally specific and concrete. As Honold [8] observed, cultural orientation manifests itself in artifacts (technical products) and institutions (organization). The female, elderly Indian user of German washing machines who meets a male, young usability professional from Germany may not reveal the

subtleties of her preferences for quick wash programs (she wants morning tasks to be finished by noon (app. 3 hours to do the work), top-served machines (she can control the water level in an environment with less water), and similar highly contextual issues. In relation to usability, age and gender issues have to be considered together with the objective of users, environment, infrastructure, division of labor, organization of work, mental models based on previous experience and tools to understand the nature of the usability problem list [8]. The users' cultural background may be even more complicated than age and gender issues suggest. One example is when the availability of Arabic language interface in Gmail ™ gives an bicultural (north African and Danish) middle-aged occasional user a good feeling, even if he does not at the moment apply the interface to write an Arabic mail to some of his Arabic speaking friends, and even if the possibility of writing in Arabic is not accessible – just because the availability of an Arabic interface to the email system indicates the possibility of accommodating Arabic email dialogs. His daughter, who is a power user of Gmail, may however, experience that when she changes from Danish left-to-right to Arabic right-to-left language interface the functions switch place and she is reduced to a novice user. Switching to a culturally different interface may ruin a user's memorized information structure.

We offer the following recommendations in order to avoid producing biased lists of usability problems when doing user tests, especially cross cultural tests:

1. Balance out potential "hidden user groups" within user segments, for example, users who adapt quickly to international test conditions (used to foreigners) vs users who do not (are not used to foreigners), and culturally sensitive (traditional, rural) vs not sensitive (modern, urban) users.
2. Calculate the detection rate for each "hidden user group" [10]. To avoid missing critical usability problems, pick evaluators from an evaluator group suitable to the "hidden user groups" (calculate the evaluator effect).
3. Have different versions of your test protocol ready that include different types of scenarios, such as Bollywood dramatic (India closed users) vs traditional (China, Denmark, India open users), and/or different probing questions such as direct and frank (Chinese) vs. indirect (Denmark, India).
4. When writing up the report, have different templates for different clients like foreign clients vs. home market clients.
5. Plan to repeat tests in China after 5 years, because target users change very quickly, although this may not be necessarily true in India or Denmark.

Acknowledgments. This study was co-funded by the Danish Council for Independent Research (DCIR) through its support of the Cultural Usability project. A big thank to Thomas Plocher, Honeywell, and Apala Chavan, Human Factors International India, and to the others who gave us access and helped us in the companies.

References

1. Arcury, T.A, Quandt, S.A.: Participant Recruitment For Qualitative Research: A Site-Based Approach To Community Research In Complex Societies. human Organization 58(2), 128–133 (1999)
2. Barnum, C.M.: Usability testing and research. Longman, New York (2002)

3. Burawoy, M.: Global Etnography. California Press (2000)
4. Clemmensen, T., Plocher, T.: The Cultural Usability Project (CULTUSAB): Studies of Cultural Models in Psychological Usability Evaluation Methods, Invited contribution to a parallel session. In: HCI International, (Beijing, July 25-27, 2007)
5. Ericsson, K.A., Simon, H.A.: Protocol Analysis. Verbal reports as data. Cambridge Massachusetts (1993)
6. German, T.P., Barrett, H.C.: Functional Fixedness in a Technologically Sparse Culture. Psychological Science 16(1), 1–5 (2006)
7. Glaser, B.G.w.t.a.o.J.H.: Remodeling Grounded theory [80 paragraphs] Forum Qualitative Sozialforschung / Forum: Qualitative Social reseach [Online Journal] (March 4, 2004)
8. Honold, P.: Cultural and context: an empirical study for the development of a framework for the elicitation of cultural influence in product usage. International Journal of Human-Computer Interaction 12(3&4), 327–345 (2000)
9. Kjeldskov, J., Skov, M.B., Als, B.S., Hoegh, R.T.: Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In: Brewster, S., Dunlop, M.D. (eds.) MobileHCI 2004. LNCS, vol. 3160, pp. 61–73. Springer, Heidelberg (2004)
10. Law, E.L.-C., Hvanneberg, E.T.: Analysis of Combinatorial User Effects in International Usability Tests. in CHI (Vienna, Austria 2004), pp. 9–16 (2004)
11. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, New York (2007)
12. Rau, P.-L.P., Choong, Y.-Y., Salvendy, G.: A cross cultural study on knowledge representation and structure in human computer interfaces. International Journal of Industrial Ergonomics 34(2), 117 (2004)
13. Trost, J.E.: Statistically nonrepresentative Stratified Sampling: A Sampling Technique for Qualitative Studies. Qualitative Sociology 9(1), 54–57 (1986)

Getting the Most Out of Personas for Product Usability Enhancements

Jianming Dong, Kuldeep Kelkar, and Kelly Braun

PayPal, 2211 North 1st Street, San Jose, CA
{jidong,kkelkar,kbraun}@Paypal.com

1 Background of Personas and Its Challenge in Driving Usability Improvements

In User-Centered Design, there is always the need to precisely define the user attributes, so that the product can be designed based on the patterns of these attributes. The methods of user definitions include quantitative segmentation analysis, as well as qualitative research on the patterns of user behaviors. Because these attributes are often complicated and abstract, researchers may have difficulty communicating these patterns to other team members.

The concept of creating personas came into play as one solution to the communication problem above. One typical definition of persona (from usability.com) is "a description of a specific person who is a target user of a system being designed, providing demographic information, needs, preferences, biographical information, and a photo or illustration". Multiple personas are often developed for various stages of design to represent the spectrum of the target audience. User scenarios often use personas to represent the subjects who will interact with the system being designed.

User research practitioners have developed processes and practices in the area of personas over several years (Cooper 1999, Grudin, and Pruitt, 2002, Pruitt and Adlin, 2006). This method has also been applied to broader research areas, such as needs analysis, task analysis, and market research. With the aid of personas, many user researchers and product stakeholders have been able to communicate user needs and behaviors more effectively.

Despite of broad use of personas to support many decision making process in companies, various skeptics remain (Chapman and Milham: 2006). Some of the common skepticisms include:

- Not fully believing in the value of personas.
- Not knowing how to develop valid personas.
- Challenges aggregating knowledge across multiple user studies to build personas.
- Not knowing how to make use of personas after they are developed.
- Challenges of evangelizing research findings to non-research organizations.

The PayPal User Research team developed a set of personas for its target audience over a series of collaborations. The personas have been communicated to entire organizations to help support their decision making processes. Personas have also provided a solid starting ground for several other research methods.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 291–296, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 Process of Creating Personas (Identifying and Prioritizing Attributes)

Personas need to be based on concrete qualitative and quantitative user research. The PayPal User Research team started with a clear goal, a phased methodology, and defined team members and roles for creating their personas. Each stage was documented and reviewed with management throughout the process.

The first step was to aggregate all existing knowledge using data and research reports from previous user research, market research and business unit efforts. The core team included several researchers who spent hours with actual users from the core user segments. The core team possessed a wide range of domain and user segment knowledge. Periodic reviews with cross-functional team members including marketing and business units proved valuable.

Tip: Start with clearly defined goals and a cross functional team.

Tip: Aggregate existing knowledge and research reports before starting the Persona development process.

The second step included defining key attributes (or "variables") that were based on existing research. Variables were defined as key attributes that defined our user population. One example of a variable in the PayPal Merchant Services business would be 'annual online revenue'. Each variable has some sort of 'scale'. For the above example the scale is: Annual Online Review 'Less than 100K, 100K to 250K, 250K to 1 Million, 1 to 5 million, Greater than 5 million'.

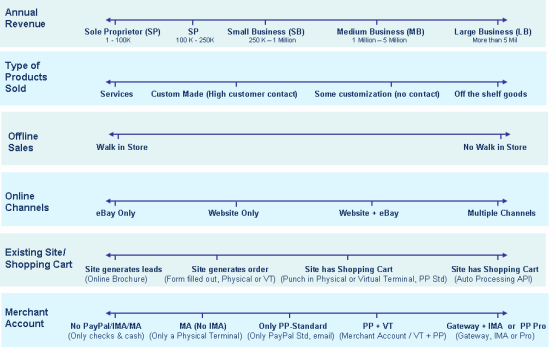


Fig. 1. Key variables, each with a scale

At PayPal we have used several of these variables for previous participant recruiting for different type of research. We already had a great deal of domain knowledge when we started this project. We had several hypotheses, many of which had been validated during previous field visits, usability tests and focus groups.

Tip: Start with domain and user segment knowledge derived from research.

The next step included a collaborative exercise using a large room with big walls, easel pads, and different color pens. This exercise included defining all 'key'

variables based on existing qualitative and quantitative research. Each variable was written down and its scale discussed. We decided to us about 35 variables. There were many other variables but not all were deemed as 'key'. For example, 'number of children' could have been an attribute but it was not deemed as a 'key' variable for our user segment(s). Of course our personas might or might not have children, but that was not a major contributing factor that defined how they would use PayPal to receive payments as part of running their business.

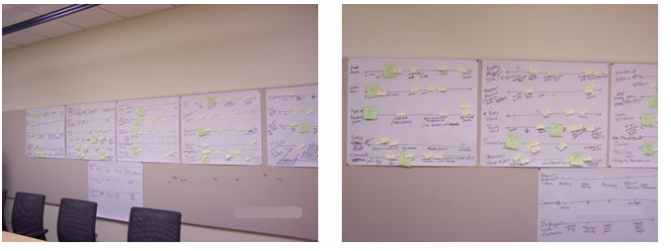


Fig. 2. Variables and manual clustering exercise

Tip: Use easel pads and a big wall or a very large white board.

Once we had all key variables defined we used Post-it ® notes to plot participants we had visited or interviewed. We developed hypotheses and started creating manual clusters for some of those key variables. During this 'manual clustering' session, Post it notes let us be flexible because we could easily move them around.

Out of this manual clustering exercise Personas were born. Then the next step included defining supporting attributes, personalizing personas, assigning names, business details, quotes from actual merchant visits and several other details.

Tip: Do not start the project thinking you will define 'x' number of personas. Let the clustering define how many personas you will have.

The next step included validating the personas. At PayPal we invite participants regularly for different types of research for various projects. This was a great opportunity to validate our personas. For each participant, the researcher would use a standard set of questions to get information on all persona variables. This process helped us validate our persona set.

As part of the project, we decided to document personas in different formats for different types of consumption. In the business world there are always competing things that need your attention. There is always a need to document and present your work in smaller bits which are easy to consume, get people interested and then provide them detailed versions.

We created the following documents:

- Brochures (One page brochures)
- Foundation Document (Word document with all details)
- Posters (Large posters for hallways, kitchen areas)
- Presentation (PowerPoint deck for presentation)
- Personas Intranet Site

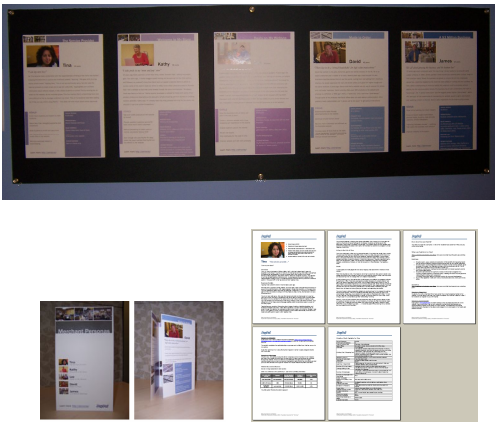


Fig. 3. Persona documents and presentations

Tip: Brochures were a big hit. They are widely used because they are easy to carry and handle.

The Brochure (single 14x8 inch folded page) was a big hit because they described all personas on one page which was easy to carry and handle. We see team members keep these on their desks and carry them to meetings. The Brochure defines all personas on one page front and back, has photos, name, and key attributes; has some description and defines key differentiating variables.

Posters (several 11x17 inch paper) printed on large glossy paper are used to attract attention and as marketing material. We see employees sipping coffee and reading details of personas.

For anyone interested in using details of the personas we developed a detailed 'foundation document'. This foundation document included details of each persona, a day in their life, key tasks, tools used to perform these tasks among other things. We also have key usage scenarios, perceptions and future feature enhancement sections for everyone to contribute to. The foundation document includes key market research segment based information, details of all variables and ways to find more information.

The Personas intranet site helped us communicate and deliver all documents to a very large audience.

Creating personas which are based on research is only a part of building successful personas. The key is to involve teams during its development, review with key stakeholders throughout its process and launch with smaller teams (beta release). Personas are successful only if they are used by different teams in meaningful ways.

3 Using Persona in Design Iterations and Usability Evaluations

Once personas were defined they were released to a group of UI designers and a single marketing team to get early feedback and see how they would use Personas. This served as a 'beta' release, helping us with further refinements, understanding

how other teams might actually use it. It helped us develop some 'success case studies' of answer other teams questions on 'how to use personas'.

Tip: Have a 'beta release' and develop 'success case studies' before releasing to a wider audience.

Personas have been widely used by PayPal teams. Personas are being used to define user needs, communicate user wants, develop scenarios and storyboards, create mood boards for visual styles, define page level content requirements as well as recruit participants for research.

Some design teams have used Personas to present their designs during executive reviews. They have used a persona to narrate existing pain points and then explain how this persona will be using the new product and its user interface. People 'get it' when there is a real person (persona), with a real problem and a product that solves that unmet need.

User Research teams use personas to achieve consistencies across studies and defining a 'user' for recruiting participants. Personas help create meaningful 'tasks' which are representative of each persona. Research teams continue to validate user behavior using on going research.

4 Communicating Persona to the Business

The personas are of little value if they were not effectively communicated to various organizations throughout the company. To maximize the benefits of personas, the User Research team made significant efforts to evangelize the deliverables of the personas.

First of all, the User Research team presented not only the results of the personas, but also the rationales, and example of application of the results in different areas. Through these presentations, the audience was able to relate their projects with the personas, and thus convinced by its value. After these presentations, they were generally motivated to use it in their decision-making process. The business owners also helped evangelize this research by exhibit posters and distribute the brochures in their respective organizations.

Secondly, the User Research team also used one of the on-going projects (bulk shipping) to test the real use of the personas. This exercise provided a comprehensive show case for the actual use of personas to the business. Also, it provided many interest insights to enrich the understanding of each persona.

5 Learning / Next Steps

After the personas were completed, the User Research team found that not all the personas receive the same level of attention. Personas representing larger business values generally received more interests, and thus were discussed far more extensively than the other personas. So we learned that a persona valuation is a necessary step of the process. Highly important personas often need to be created based on more extensive research.

We also found that the user segmentation research which Market research team conducted was very relevant to the persona research. Although user segmentation study and persona research yielded different user classifications, the concepts actually converge very well. These different classifications looked at user groups from different angles, and thus can be used for different purposes.

References

1. Cooper, A.: The Inmates are Running the Asylum. Macmillan, New York (1999)
2. Grudin, J., Pruitt, J.: Personas, participatory design and product development: an infrastructure for engagement. Paper presented at Participatory Design Conference 2002. Malmo, Sweden. Online (June 2006) at: http://www.research.microsoftcom/research/coet/Grudin/Personas/Grudin-Pruitt.doc
3. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Morgan Kaufmann, San Francisco (2006)
4. Chapman, C., Milham, R.: The Persona's New Clothes: Methodological and Practical Arguments Against a Popular Method, HFES 2006 Proceedings, San Francisco, CA

Testing Object Management (TOM): A Prototype for Usability Knowledge Management in Global Software

Ian Douglas

College of Information, Florida State University
Tallahassee, Florida
idouglas@lsi.fsu.edu

Abstract. The collection and sharing of results from usability laboratories around the world has not yet made good use of emerging models of Internet-based knowledge sharing technologies. This paper will present a model for a system that could improve the sharing of knowledge on a global scale and also facilitate the linkage of design guidelines and patterns to the accumulated evidence from the many worldwide studies that are not processed into academic publications.

Keywords: knowledge management, usability testing, global software development.

1 Introduction

A number of criticisms have been directed at lab-based usability testing. These are mainly centered on the cost of setting up and running the labs and the effect they have on users' behavior. Despite such criticisms, a number of labs have arisen all over the world that are constantly conducting tests and obtaining potentially useful information that could be implemented to support usability design. The international spread of usability expertise is increasing greatly [15].

A criticism that is not often made about such labs is that the data they produce at great expense is not cost-effectively used. What is learned from usability tests is passed on to others generally through non-standard documents with a limited distribution. Results from usability laboratories are sometimes presented within papers published journal and conferences. However, the results presented in such papers are not readily available in a standard digital format that would allow cross study, collection and comparison of the data. There have been attempts to draw from such reported studies in order to provide some support for commonly proposed design guidelines. For example, http://www.usability.gov, managed by the US Department of Health and Human Services, references academic papers that provide some supporting evidence to its published guidelines. This activity has proved surprisingly difficult, with most guidelines having at most only a few studies providing support.

A problem in using academic papers is that they are not often focused on a particular design issue, and they are formatted in a way that suits the needs of the academic community rather than those of practitioners trying to identify specific evidence-based design guidelines. Usability tests on particular software are more

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 297–305, 2007.
© Springer-Verlag Berlin Heidelberg 2007

directly focused on design issues, but the report formats are often varied and there are no established publishing outlets. There has been an attempt to address this problem of reporting standards by the introduction of the standard document format for usability test reporting (the common industry format). While this goes part of the way to solving the problem outlined above, the objective of this paper is to propose an alternative scheme that provides a more complete solution and allows knowledge about cultural differences in design to be more clearly identified.

1.1 Current Approach

Currently, Microsoft Word or Adobe PDF documents are the primary methods for communicating information related to usability problems. Finding a document related to a specific usability problem is often difficult, and even if such a document is found, locating specific information about the problem within the document is also difficult. Navigation within the document is complicated by the fact that it is not organized according to a standard format. Creating a standard document format is the advocated approach to solving this problem, and this approach has become popular in some domains, such as the reporting of clinical trials. The Consolidated Standard of Reporting Trials, or CONSORT [3], establishes guidelines for the standard reporting of clinical trials. These guidelines help researchers better communicate the results of their trials so that they can be more clearly understood, and can be compared and consolidated. The National Institute of Standards and Technology has established the Common Industry Format (CIF) [11] to achieve the same goal for usability testing (ANSI INCITS 354-2001).

Establishing common document formats is only a partial solution, however, because the flow of usability information is still constrained by the traditional document-centric view of knowledge communication. If someone is interested in knowing what test results relate to a particular technology or design attribute, he or she has to identify the appropriate documents and process the contents before collating all the related results.

1.2 Network/Object Centered Approach

A number of domains (e.g. software engineering, learning, technical documentation) have accepted that one way to increase the reuse and sharing of knowledge is to granularize a system into objects and to categorize and store the objects in digital libraries that can be browsed and searched. The aerospace and defense industries association of Europe has developed S1000D as an object based approach to the construction of technical documentation (see http://www.s1000d.org/).

An example of a networked environment to support the reuse and sharing of knowledge objects which are normally integrated in reports is Net-Centric Performance Improvement (Net-PI) [5] [14]. Net-PI is a system that allows teams to collect and critique knowledge during a human performance analysis project. The resulting knowledge is compiled into both a conventional report and a set of XML-based information objects that capture the key analysis data. These objects are included in a digital library that allows subsequent analysts to build on existing knowledge, rather than having to reanalyze the problem from scratch.

The general idea of using the power of the Internet to increase the generation, sharing and dissemination of new knowledge is gradually developing. Initially the document focused approach people were already familiar with was merely translated to the Internet domain in the form of e-journals. The newest development is the concept of cyberinfrastructure, which strives to use the power of Internet technology to more effectively promote progress in science and engineering [2]. A key part of this is improving the processes of collaboration and knowledge sharing.

2 Usability Test Result Object

There is scope for adapting object-based standardization efforts in other domains, particularly in e-learning, to facilitate the reuse and sharing of knowledge in usability. A starting point involves the capture of key data from tests in XML-based digital objects. These objects can then be catalogued according to user profiles and goals (or use scenarios) in a globally accessible test knowledge repository. Usability test objects have the advantage of being both embeddable within a traditional document and storable in digital repositories where they can be accessed, collected, and compared by a global community of usability specialists. An initial standard for a test object schema is based on key fields derived from the common industry format (Table 1).

Table 1. Initial data elements for a usability test object

Element	Description
The use case/ test scenario	One specific goal assigned to a test subject
Specific interface component	The design components included in the test, e.g. date selector, form, shopping basket
Subjects	Details on the subjects used in the test
Intended end-user	Description of the intended end user
Test method	How the test was conducted
Problems identified	Specific elements of the design that caused the users' problems and why
Recommended solutions	How to address each problem noted above
Dublin core metadata	This is the accepted standard for web resource meta-data and includes such items as author and date created.
Unique object identifier	This would allow registration of the object in a digital library

For each design, a digital test results object (TRO) is collected, which can be stored in a database or a single XML file. The TRO can be compiled into a conventional report; it can also be registered in a digital library of objects relating to the goal. A range of data would be collected in the TRO; an initial specification for the contents is presented in Table 1. As with all standards, different parties need to comment on the proposed elements before the standard can be fully adopted. It is also possible to allow extensions to core standards or to make some elements optional, depending on the circumstances.

3 Tools for Knowledge Sharing

There is a need for a set of tools to assist in the Test Object creation and access. Test Object Management, (TOM, see figure 1) is a prototype for such a tool. Existing user test data from third party usability studies is used to illustrate that objects can be derived from existing CIF compliant reports. TOM allows the creation of objects and the storage in a digital repository that is catalogued according to software types, user roles and user goals.

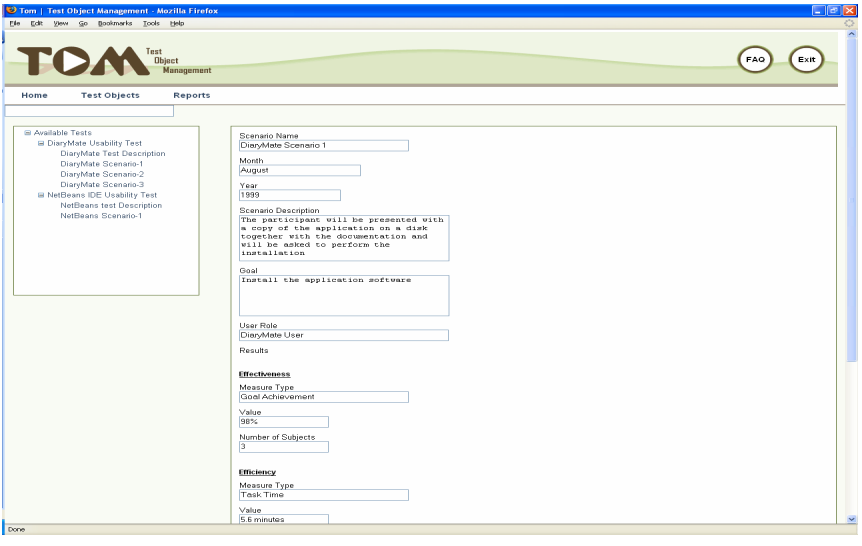


Fig. 1. A screen from the Test Object Management system, which displays test projects and scenarios, and provides test object editing

Under the TOM system, labs will separate the collected data not just by the product tested but also by the specific users' goals in the interaction scenarios tested (e.g., booking a flight). The specific interface design components relating to the goal will then be identified for different interfaces designed to assist users in achieving this goal. One goal may have several different interface design components or specific designs for specific regions or countries.

TOM is just one example of a tool for the creation and management of test result objects. It is possible to adapt other web-based meta-data creation tools with templates for test result objects [9]. Another possible approach is to create add-ons and templates for popular word processing tools. This would allow for the creation of reports with tools users are already familiar with, while enabling the extraction of data to be converted into the object format.

4 Repository Model

A federated scheme similar to that of the latest thinking in repositories for digital content used in training is a useful model for usability testing. Kahn and Wilensky [8]

conducted some of the earliest work in confederated libraries and the design and development of infrastructures for open architecture. Further work has been conducted by the Digital Library Research Group at Cornell University and the Corporation for National Research Initiatives (CNRI) [12]. However, the most developed model for this approach is the Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA), which is a collaborative activity emerging from the U.S. Department of Defense's Advanced Distributed Learning (ADL) initiative [13].

CORDRA was created with the assistance of the Corporation for National Research Initiatives (CNRI) and Carnegie Mellon University's Learning Systems Architecture Laboratory (LSAL). CORDRA is a model for a global infrastructure for the federation of content repositories. Although primarily targeted at e-learning content, it is flexible enough to incorporate many types of content objects. CORDRA allows groups of separate repositories to form federations by entering their content in a central registry. Registers can themselves be registered in a higher-level Registry of Registries, with one root-level Master Registry of Registries where a unique identifier is assigned to each registered item. The individual federations can vary with regards to metadata standards, access policies, and organizational principles. Applied to usability testing, such a scheme would allow each individual testing organization to operate its own repository and to decide what it shares through a central registry. There can be registries for each country that all feed into the root-level registry.

4.1 Repository Tools

In addition to standards for the Test Result Object and the tools to create objects, standards must be created for the digital library or repository through which the objects would be catalogued and accessed. There are already well-developed models in other domains (e.g. learning content objects) that can be readily adapted for the purpose of TROs. (See for example the sharable content object reference model, SCORM at www.adlnet.org.) The main tool needed is one for organizing groups of objects into meaningful collections.

One possible scheme we will investigate is organizing objects by identified user goals. In this scheme, each separate test on one user goal results in a TRO, which is in effect a specific instance of a general class of tests for a specific type of interface design component. When tests conducted in different countries reveal similar results, they are included as an object that is linked to the main class of tests. When tests in different countries reveal different results, a new subclass for that country is created.

Another possible approach is social book marking [7]. Social book marking works by utilizing the contributions of the diverse, independent, and decentralized contributions of users through the tagging of content. This is essentially the approach used by websites such as flicker.com and del.icio.us. Such an approach would enable the community of users (educators, testers, developers, and researchers) to identify and classify what is important to them.

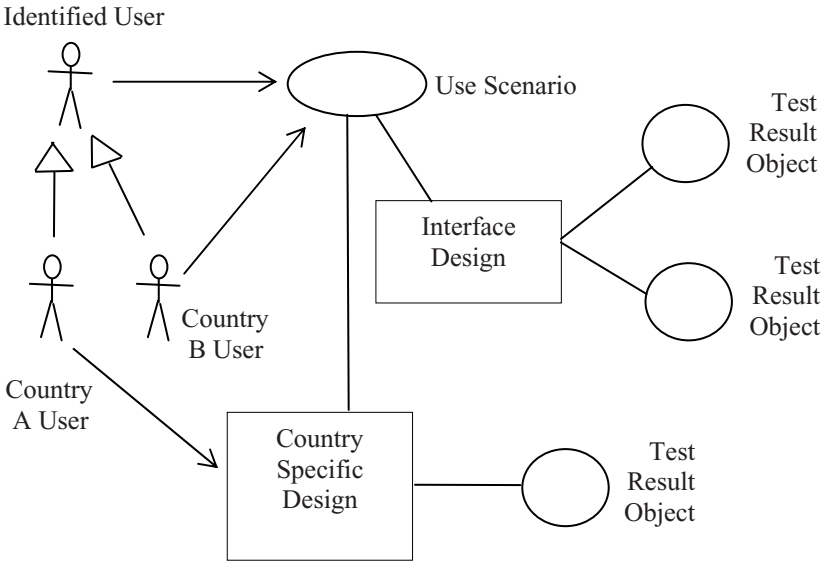


Fig. 2. An organization scheme for linking usability test objects to country specific interface designs for user roles and goals

4.2 Connecting Guidelines and Patterns

Effectively collecting data derived from international tests is a first step towards more efficient test reporting and archiving. The next step would be to use the repository to investigate how the stored data can translate into effective design advice for developers of new technology. This is necessary to analyze the data and derive effective guidelines for developers. The system proposed would enable the aggregation of data from a number of labs and tests to provide a level of confidence in backing certain recommendations (see Figure 3). An effective system would reduce usability errors more effectively by preventing them from arising in the first place.

The National Cancer Institute supported an effort to collect web design guidelines and to identify the evidence that supports each guideline [10], relying mainly on work published in academic outlets. Many tests documented in usability labs are not available because there is no place to publish the results and non-academics do not always have the time to write up the results in a form acceptable for an academic publication. An open Internet-based repository in which any lab can publish results has the potential to accumulate more substantial evidence for certain guidelines and to provide more specific indications of how guidelines can be implemented in certain types of interfaces, or for users in certain countries. Studies in cross-cultural usability suggest a need to make distinctions in certain guidelines and priorities [4] [6] [17].

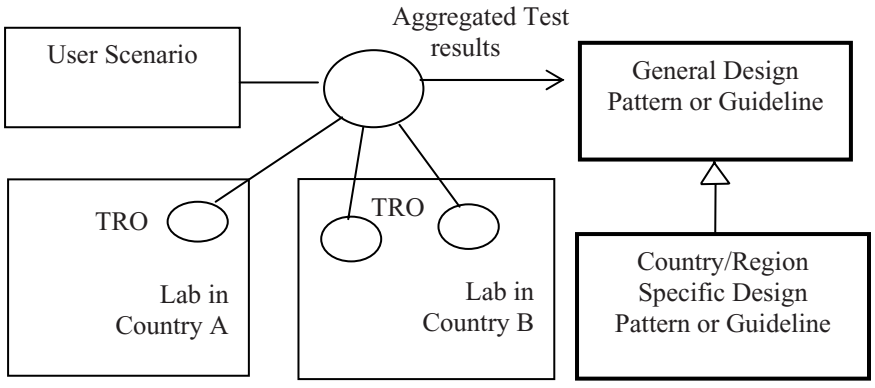


Fig. 3. An organization scheme for linking usability test objects to country specific interface designs for user role and goals

The design patterns movement prevalent in a number of design domains presents a mechanism whereby internationally collected testing data can be abstracted into design patterns that have variations for particular regions or countries. Design patterns arose in architecture when Alexander [1] noted that certain abstract patterns could be identified across a range of successful room, building, and city designs. Alexander determined a patterns language whereby these patterns could be described in a standard way and then rated and discussed by the design community. There have already been attempts to develop a set of patterns for interface design [16] but, as with the National Cancer Institute guidelines, results reported in a few academic papers are mainly used to validate the patterns. Patterns provide a potentially useful mechanism to aggregate results from a range of tests.

Another problem is that of context. An interface that has been proved to support a user goal in one context may not work well in another. This problem also applies to the concept of patterns, where abstracting problem-solution pairings from a limited number of concrete problem analyses may give a false indication of the general applicability of the pattern. In a sense, this perceived limitation derives from seeing both the object and pattern creation activities as discrete, rather than continuous processes. As more people share their test data and identify different contexts where the perceived patterns and guidelines do not apply, the community can determine when new or special cases, guidelines, and patterns are needed. A system like the one proposed in the previous section could provide the data for the community to refine and recognize the limits of patterns and guidelines.

5 Conclusion

This paper proposes to investigate an open-standard based approach to the sharing of test results in the form of digital objects. Not only would such an approach reduce the needless replication of tests that occurs when there are no public records of previous tests conducted, but it would also allow for the accumulation of a good deal of evidence to support certain usability design patterns and guidelines. Results from labs

in different countries would facilitate the identification of cases where specific guidelines and patterns require special variations for audiences in certain countries. The emergence of such a collection of knowledge would be of great benefit to all developers. Without it, developers are either restricted in their access to usability expertise or have to rely on their own best estimate as to what interface design works best and how it should be adapted for different countries.

The proposed scheme is adapted from similar efforts to facilitate the reuse and sharing of knowledge in other domains, particularly in e-learning. It involves the capture of key data from tests in XML-based digital objects and unique identifiers for cataloguing. These objects are organized in digital repositories according to user profiles and user goals in a globally accessible repository. Usability test objects have the advantage of being both embeddable within traditional documents and storable in digital libraries where a global community of usability specialists can access, collect, and compare them. The objects can be linked to existing design guidelines published on the web. The test data objects would thus serve to produce individual reports and to accumulate a body of evidence to support both general and specific design guidelines.

A global repository of test knowledge can be developed and used to help better validate design guidelines and patterns. The proposed system would allow this knowledge to be aggregated and distributed on a global scale, and it would be useful in identifying regionally and culturally specific variations in design guidelines.

References

1. Alexander, C.: The Timeless Way of Building. Oxford University Press, New York (1979)
2. Atkins, D.E., Droegemeier, K.K. et al.: Revolutionizing Science and Engineering Through Cyberinfrastructure (2003) Retrieved June 9, 2006 from: http://www.communitytechnology.org/nsf_ci_report/
3. Begg, C., Cho, M., Eastwood, S.: Improving the Quality of Reporting of Randomized Controlled Trials: the CONSORT Statement. Journal of the American Medical Association, pp. 276–637 (1996)
4. Del Galdo, E., Nielsen, J.: International User Interfaces. John Wiley & Sons, New York (1996)
5. Douglas, I., Butler, J., Nowicki, C., Schaffer, S.: Web-Based Collaborative Analysis Reuse and Sharing of Human Performance Knowledge. In: Proceedings of the Inter-Service/Industry Training, Simulation and Education Conference (I/ITSEC), pp. 1023–1030 (2003)
6. Fernandes, T.: Global Interface Design. Morgan Kaufmann, San Francisco (1995)
7. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social Bookmarking Tools: A General Review. D-Lib Magazine 11 (2005)
8. Kahn, R., Wilensky, R.A.: Framework for Distributed Digital Object Services. International Journal on Digital Libraries 6, 115–123 (1995)
9. Malaxa, V., Douglas, I.: A Framework for Metadata Creation Tools. Interdisciplinary Journal of Knowledge and Learning Objects 1, 151–162 (2005)
10. National Cancer Institute: Research-Based Web Design and Usability Design Guidelines (2006) Retrieved November 29, 2006 from: http://www.usability.gov/guidelines/

11. National Institute of Standards and Technology: Common Industry Format for Usability Test Reports. Industry Usability Reporting Project (2001) Retrieved November 30, 2006 from: http://zing.ncsl.nist.gov/iusr/documents/cifv1.1b.htm
12. Payette, S., Blanchi, C., Lagoze, C., Overly, E.A.: Interoperability for Digital Objects and Repositories: The Cornell/CNRI Experiments. D-Lib Magazine (1999)
13. Rehak, D.R., Dodds, P., Lannom, P.: A Model and Infrastructure for Federated Learning Content Repositories. In: Proceedings of the Interoperability of Web-Based Educational Systems Workshop, 14th International World Wide Web Conference (2005)
14. Sasson, J., Douglas, I.: A Conceptual Integration of Performance Analysis, Knowledge Management and Technology: From Concept to Prototype. The. Journal of Knowledge Management 10, 81–99 (2006)
15. Schaffer, E.: Offshore Usability: Helping Meet the Global Demand? Interactions XIII.2 (2006)
16. Tidwell, J.: Designing Interfaces: Patterns for Effective Interaction Design. O'Reilly Media, Sebastapol, CA (2005)
17. Vatrapu, R.: Culture and International Usability Testing: The Effects of Culture in Interviews (2006). Retrieved November 9, 2006 from: http://scholar.lib.vt.edu/theses/available/etd-09132002-083026/unrestricted/Vatrapu_Thesis.pdf

Assessing Usability Problems in Latin-American Academic Webpages with Cognitive Walkthroughs and Datamining Techniques

María Paula González[1,2], Jesús Lorés*, and Antoni Granollers[1]

[1] Grup GRIHO – Universitat de Lleida - c/Jaume II, 69 –25001 Lérida, Spain
{mpg,tonig}@diei.udl.es
[2] Dept. of Computer Science and Engineering – Universidad Nacional del Sur
Av. Alem 1253 – 8000 Bahía Blanca, Argentina
mpg@cs.uns.edu.ar

Abstract. Qualitative usability evaluation is usually included within the Evaluation Stage in Usability Engineering through a *Qualitative Usability Testing process QUT*. This *QUT* process includes the application of methods that have been defined focusing on the evaluation of a particular interactive system, becoming highly expensive when a context of use has to be evaluated (by analyzing a large number of interfaces belonging to that context) in order to detect common usability problems from a qualitative viewpoint. This paper presents the $QUTC_{KDD}$ methodology which incorporate techniques from Knowledge Discovery in Databases, extending the existing *QUT* process in order to solve the above situation. To illustrate the $QUTC_{KDD}$ methodology, an experimentation related to a particular Latin-American context of use is also discussed.

1 Introduction and Motivations

Usability is a software attribute associated with the "ease of use and to learn" of a given interactive system [7]. Nowadays usability evaluation is becoming an important part of software development based on Usability Engineering (*UE*) [11,14]. During usability evaluation a *qualitative usability estimation* is performed, since no quantitative measure can be expressive enough to represent something so complex as a usability problem [4,7]. Indeed, qualitative results are collected through a *Qualitative Usability Testing* process (*QUT*) which includes a number of different methods (such as Cognitive Walkthrough (CW) [21], etc.) focused on analyzing the interface of a particular system [2,7,13,18,20]. In spite of the differences present among these methods, there exist a number of common elements (see dashed box in Fig. 1) that can be identified in any *QUT* process: an *interactive system S* which is being assessed using the *QUT* process; a *generic evaluation method M* for qualitative usability testing (selected among the possible methods mentioned above); a *particularization* (or instance) M' of the method M, defined for applying M to the particular usability evaluation of

* In memoriam.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 306–316, 2007.
© Springer-Verlag Berlin Heidelberg 2007

the system S; a dataset representing the *results* obtained from the application of M' over S; a number of *visual representations* (visualization) of the above results which facilitate their processing and analysis; a final *usability report* containing the results coming from the whole qualitative usability evaluation process. Besides, some typical steps are carried out when applying a QUT process to a generic system S, namely:[1]

- *Planning.* A particular method M is selected by the evaluation team, depending on several aspects (e.g., its adequacy to the system S at issue). According to M, sometimes a redefinition of the team should be done. Besides, the conditions to perform M (timetable, physical environment, equipment, etc.) should be stated. Note that most QUT methods are defined in a general way. Consequently, many times a specialization M' of M should be defined in order to adequate M to the particular process in progress. The definition of M' includes the selection of appropriate ideal values for the context of use involved, and a more specific definition of the main features of M (e.g., the definition of tasks in CW). Additionally, the selection of specialized software may be required.
- *Execution.* This step materializes the real execution of M', which may require the use of a specialized software with adequate parameters. The members of the evaluation team will apply M' to test the system S under evaluation, carrying out actions such as task completion, time capture, collecting usability data, etc. Depending on the features of M', the process will be carried out different numbers of times. It must be stressed that due to the nature of QUT methods, qualitative information is collected and complemented with natural language explanations.
- *Processing and analyzing results.* A crucial part of the QUT process is the processing and the analysis of the obtained results. In the worst case, the final usability problem set is created just by compiling the output data from M' without any other consideration. In an ideal situation, the above output data are processed and analyzed to achieve a general diagnosis of the usability problems in S. To do this, the members of the evaluation team should compare those data in order to detect agreements and disagreements between the different applications of M' during the previous QUT step. Any usability problem should be carefully examined to decide if its relevance is strong enough. Frequently, this examination implies the consideration of natural language explanations that complement the alphanumeric data available. The use of different kinds of graphical representations (e.g., diagrams) can enhance the visualization and processing of the obtained results.
- *Reporting conclusions.* As a final output, the QUT process returns a set of usability problems detected in the system S under evaluation. To report this set there are different standards and reporting sheets [7].

Most QUT methods are focused on analyzing the interface of a particular system S, becoming complex when a large number of real cases from the same

[1] We are aware that in actual software development projects under *UE* some parts of these steps might overlap.

context of use have to be jointly considered to provide a general diagnosis, since a considerable amount of qualitative information must be treated simultaneously. In this situation, it would be desirable that the evaluation team be able to consider all such information with a reasonable cost and without losing the richness of the qualitative perspective. Note that identifying the most general usability problem patterns of a context of use from a qualitative viewpoint can help not only to evaluate a new interface belonging to it, but also to prevent usability errors when a novel interactive system is being developed.

To cope with the above problem, this paper presents a new methodology called $QUTC_{KDD}$ in which association rules (a technique from *Knowledge Discovery in Databases* (*KDD*)) are used to extend the existing *QUT* process in order to provide a general usability diagnosis of a given context of use C. Starting from the data obtained by applying a traditional *QUT* method to a sample group of real interactive systems from C, the goal is to automatically detect qualitative usability problem patterns, each of them describing a relevant usability feature of C from a qualitative viewpoint.

2 Knowledge Discovering in Databases: Overview

KDD is a discipline which involves "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data". This extraction can be modelled as a process in which different steps [10,16] can be identified: 1) *Data preprocessing:* This first step condenses a number of intermediate steps, such as data cleaning, integration, selection and transformation. Data cleaning routines attempt to fill in missing values, smooth out noise, and correct inconsistencies in the data. If required, multiple data sources may be combined in a single database, and additional selection and transformation operations may be performed to unify and consolidate data formats. 2) *Datamining (DM):* Once data has been preprocessed and condensed in a database, datamining techniques can be applied in order to extract relevant patterns from it. There exist several datamining techniques, such as association rules, decision trees, etc. (for a detailed discussion see [10,15]). 3) *Pattern evaluation and knowledge representation:* Different criteria (e.g. interestingness measures) are applied to evaluate the patterns obtained as an output of the datamining algorithms. 4) *Visualization:* Finally, visualization and representation techniques are used to present the mined knowledge to the user.

Based on the previous steps, a typical software platform for *KDD* would involve the following major components: a) a database or information repository, consisting of databases or spreadsheets; b) a *KDD* engine, which consists of a set of functional modules for accomplishing datamining algorithms, providing as well modules for pattern evaluation and visualization; c) a Front-End module which allows the final user to interact with the *KDD* engine, typically by means of a specialized *datamining query language* (DMQL) [10]. Different general-purpose platforms like WEKA [22] or Orange [6] have been developed to carry out the *KDD* process, each of them with their specific DMQL language.

Such query-oriented languages range from very basic command-line interpreters (as the Command-line Interface provided by the WEKA platform [22]) up to more sophisticated tools such as specialized scripting languages (as the one provided by the module `orngMySQL` in the Orange platform [6]), MSQL [12], or more recently KDDML [17], among others.

Association rule mining [10] is a powerful datamining technique which allows to find hidden relationships among attributes in a transactional database D. Every transaction consists of a set of items $I = \{i_1, \ldots, i_m\}$. An association rule (AR) is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, and $A \cap B = \emptyset$. In addition to the antecedent A (the "if" part) and the consequent B (the "then" part), an AR has several interestingness measures that express the quality of the rule. One relevant measure is called the *support* for the rule, which is simply the percentage of transactions in D that contain $A \cup B$. Other important measure is known as the *confidence* of the rule, and corresponds to the ratio of the number of transactions that include all items in the consequent as well as the antecedent (namely, the support) to the number of transactions that include all items in the antecedent. (i.e. the percentage of transactions in D containing A that also contain B). Computing association rules (ARs) is a computationally complex task, and several algorithms (e.g. Apriori or FP-Growth [10]) have been developed. The AR mining process generates usually a huge number of rules, making it necessary to provide powerful query primitives for performing selective, query based generation [12].

3 Our Proposal. The $QUTC_{KDD}$ Process

In our proposal, *KDD* is used to gather common usability features of a large group of systems $S_C = \{S_1 \ldots S_n\}$ belonging to the same context of use C in order to achieve a general diagnosis of C. In what follows we will call this new approach $QUTC_{KDD}$. By performing $QUTC_{KDD}$, the members of the evaluation team will be able to formulate *queries* to an underlying *KDD* engine, getting as an output additional, qualitative information that will condense usability problem patterns of C. The result of every query will be expressed in terms of a ranked list of *association rules* where every AR will express a hidden cause-effect relationship among usability features of S_C.

Fig. 1 shows a schematic view of the framework for $QUTC_{KDD}$, which is based on extending the traditional QUT approach by adding the following elements: a *transactional database DB* which stores all the results obtained from applying the traditional QUT process to S_C; a *KDD engine* capable of automatically computing ARs; a *DMQL Front-end* for the *KDD* engine, which should provide the interface for the evaluation team in order to pose queries (using the DMQL associated with the *KDD* platform); a *visualization module* (in most cases provided by the *KDD* engine itself) which provides different alternative graphical representations of the patterns found; and a *collection of usability problems patterns* (ranked lists of ARs) related with C. As before, we will summarize next the different steps related to applying a $QUTC_{KDD}$ process to S_C:

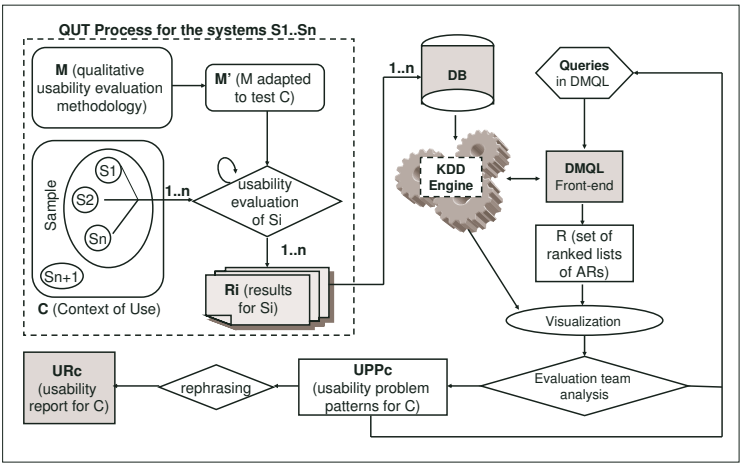


Fig. 1. Schema of the proposed Qualitative Usability Evaluation ($QUTC_{KDD}$) Process for a context of use C of interactive systems

- *Planning.* Planning the $QUTC_{KDD}$ process will involve similar steps as the ones described for the QUT process (see Section 1). Besides, a sample group of systems $S_C = \{S_1 \dots S_n\}$ must be selected to represent the context C. Choosing this sample group is not a trivial task, and the use of the selection criteria will depend on the nature of C [16]. In addition, the selection of a particular software platform and the associated DMQL must be considered.
- *Collecting the data.* Performing the method M' within all the elements in S_C is a central task. As the evaluation team carries out the usability evaluation of the elements in S_C, the obtained results (qualitative data) must be collected in an information repository (a database, a spreadsheet, a data warehouse, etc.).
- *Applying* KDD *techniques.* Once the data related with qualitative usability problems has been collected, the next step in the $QUTC_{KDD}$ process will consist in discovering patterns in that data by means of *KDD* techniques. The data repository obtained in the previous step has to be automatically transform in a transactional database [16]. The following steps will be carried out, as described in Section 2:
  - *Preprocessing the qualitative usability data.* There does not exist a fully automatic system capable to preprocess the data to clean and integrate them [16]. In the case of the $QUTC_{KDD}$ process cleaning and integrating the data is worthwhile, as the result can be used to obtain statistical information.
  - *Mining the data.* Within this step, the evaluation team can use the selected *KDD* software platform and the associated DMQL in order to pose queries using the DMQL Front-end. In a first stage, the evaluation team will post a starting query just to detect the most relevant cause-effect relationships (expressed as *ARs*) among the attributes in the database

DB on the basis of a threshold value for support, confidence and a maximum number of rules to be computed. As a result, a ranked list RAR_{i1} of *ARs* will be automatically obtained, containing the most relevant information (with possibly hidden cause-effect relationships) that needed to be considered by the evaluation team. By performing the next step (*Visualizing and analyzing* $QUTC_{KDD}$ *results*) the evaluation team will be able to analyze the relationships present in RAR_{i1}. Next, on the basis of RAR_{i1} information, more refined queries can be posted. This time, each new query for mining *ARs* will produce additional and more focused information, characterized as new ranked lists $RAR_{i2}, \ldots, RAR_{ik}$.

- *Processing and analyzing results* The use of visualization tools provided for most *KDD* platforms [6,22] can help members of the evaluation team to observe different graphics and charts representing the obtained patterns. For the elements in each ranked list, results could be displayed as a sequence of *ARs* or as a group of Cartesian charts (each of them depicting the relationships between attributes present in the *ARs* in the ranked list). The goal of the evaluation team analysis will be to decide whether every pattern found depicts a usability problem of C. Indeed, each detected pattern will be discussed on the basis of its relevance and significance. For example, the relative position of every *AR* in a ranked list expresses these features, jointly with its interestingness (support and confidence).
- *Reporting conclusions.* The $QUTC_{KDD}$ process returns as an output a set of justified usability problems patterns detected for C. each pattern has to be rephrased as a usability problem expressed in some standard format (similar to the one shown in [7]). If necessary, different visualizations of the usability problem patterns generated during the $QUTC_{KDD}$ process can be added.

Note that during the traditional Evaluation Stage in *UE* different data generated during the evaluation processes are usually compiled into databases required for the statistical analysis associated with quantitative usability evaluation. The re-utilization of these databases minimizes the cost associated with the generation of the documentation needed to carry out $QUTC_{KDD}$. Note also that the computation of *KDD* algorithms should be performed only one time to achieve satisfactory results. Finally, we would like to remark the high level of automatization present in the *KDD* stage within the $QUTC_{KDD}$ process, which is a highly desirable feature (as pointed out in [13]). In that respect the proposed approach emphasizes the role of automation, thus enhancing the predictability in the findings of this usability evaluation. However, it must be noted that the evaluation team is always in charge of controlling the $QUTC_{KDD}$ process as a whole. Even when the $QUTC_{KDD}$ methodology enhances the decision making capabilities of the evaluation team, it does not replace them.

4 Experimentation in a Latin-American Context of Use

In this Section we will summarize the results obtained after using the proposed approach to evaluate the usability of the context of use formed by academic

web sites in Spanish-speaking countries in Latin America. In particular, the $QUTC_{KDD}$ process was carried out by processing qualitative information stored in 20,400 records collected by means of seven CWs. The experimentation was carried out within the 3rd Stage of the *UsabAIPO Project*,[2] a project focused on usability research which involved the participation of more than 15 university research groups specialized in HCI, in which four categories were considered (Design Category,Content Category, Navigation Category and Search Category). The evaluation team was formed by three usability experts and two CS advanced students, all of them frequent users of the context of use under evaluation.

First, a sample group of the context of use was selected by considering the web sites of the 69 universities listed in the *Universia* portal,[3] a widely used portal about universities available for Spanish-speaking countries in Latin America. Second, CW [21] was selected as the QUT method M to be applied. Following the 2nd Stage of the *UsabAIPO Project* [9], three user profiles were considered to define M': student, professor and administrative profiles. Besides, seven CWs involving 37 tasks were defined on the basis of an exhaustive poll carried out during four months between 400 different users of academic web sites of Latin-American universities. Eight questions related with qualitative features to be tested were linked to each task (two question for each Category). Next the definition of the seven CWs and the tasks associated with the CW #1 is shown:

- CW#1 (student profile): locate and visualize the study plan corresponding to a given undergraduate degree offered by the university under study./ *CW#2* (student profile): locate and visualize information concerning academic regulations (enrollment information, etc.)/ *CW #3* (administrative profile): locate and visualize information about a training course oriented towards the administrative staff and offered by the university/ CW #4 (administrative profile): locate and visualize data for contacting a person belonging to the University administration using the option of People Search/ *CW #5* (professor profile): locate and visualize information about a posgraduate course or seminar offered by the university/ *CW #6* (all profiles): locate and visualize email access facilities/ *CW #7* (all profiles): locate and visualize a particular news made by the university.
- CW#1 (task description): *Task 1*: Visualization of types of degrees offered by the University (starting from the homepage); *Task 2*: Visualization of undergraduate degrees offered by the University (starting from the visualization of type of degree offered); *Task 3*: Visualization of study plan (starting from visualizing undergraduate degrees); *Task 4*: Short walkthrough using the study plan (courses, etc.).

Ranked values for users' cognitive effort were selected, namely (from bottom to top) *non_mesurable*, *insignificant*, *low, normal, high* and *very_high*. Besides

[2] See www.aipo.es (webpage of the Asociación Interacción Persona Ordenador – AIPO).

[3] See www.universia.es

ideal values between *insignificant* and *normal* were related to every task. The WEKA platform [22] was adopted, as it provides the implementation for mining of ARs, along with an embedded Front-End with a very simple Command-line interface for posing queries and a visualization module to improve the interpretation of final results. Next, the evaluation team was divided in two sub-teams in order to carry out two independent usability evaluation of the 69 websites considered by applying the method M' described before. As a result, each sub-team obtained 10,212 qualitative answers that were stored in two temporary spreadsheets. During the preprocessing of the data stored in the two spreadsheets, each evaluation sub-team controlled the data produced by the other sub-team, and 24 data records were rejected as they had missing values. Afterwards both spreadsheets were condensed in a final spreadsheet called `Usability_Universities` which was automatically transformed into the Attribute-Relation File Format (`arff`)[4] suitable for being processed by WEKA. Then, different queries were posted by the evaluation team to mine the data in `Usability_Universities`. General ranked list of ARs were computed using the Apriori algorithm provided by WEKA (with a support of 60%, a confidence of 90% and the selection of the best five rules obtained), each of them including in the consequent a high value associated with the general assessment of the usability problems involved in each CW. As an example, Figure 2 depicts the ranked list R#1 corresponding to the next query (TxD_y represents the question y used to test the task x of the Design Category (D)):

GetRules (Usability_Universities) where
[Consequent has {RC1_GeneralVal=very_high or RC1_GeneralVal=high}
and support > 0.8 and confidence > 0.9]}

Note that according to data in `Usability_Universities`, the information in R#1 shows: 1) the heterogeneous style in the webpages increments seriously the users' cognitive effort (AR#1), 2) the incorrect design of pull–down menus causes considerable visual noise rising the users' cognitive effort (AR#2), 3) the links' style is critic to define the users' cognitive effort (AR#3 and AR#4), 4) Information in the antecedent of AR#5 provides unbiased evidence showing that excessive animation (T1D_PageStyle=low and T1D_Animation=very_high) increases significatively users' cognitive effort and decreases heterogeneity in webpages' style. 5) In general, information in R#1 shows objectively that when values measuring users' cognitive effort increase in the last tasks of CW#1, the general assessing for CW#1 rises up to the highest possible ponderable values. Note that all the above knowledge (from 1) to 5)) was obtained from automatically mining the information stored in `Usability_Universities` by applying the APRIORI algorithm within the proposed $QUTC_{KDD}$ process.

At this stage the evaluation team was able to process the results obtained so far using the WEKA visualization module. Relationships among association rules in the ranked lists were visualized as 2-dimensional graphics. To give a formal account of the usability problem patterns that were obtained, the evaluation team rephrased them using a format similar to the one shown in Fig. 3.

[4] See http://www.cs.waikato.ac.nz/ml/weka/arff.html

```
Minimum Support: 0,5
Minimum metric (confidence): 0.9
... ... ...
Best rules found:
1. T1D_MenuDesign=normal T3D_HomogeneousStyle=high 35==>RC1_GeneralVal=very_high 35 conf:(1)
2. T1D_Links=normal T1D_PullDownMenus=normal T2D_MenuDesign=high 40 ==> T2D_VisualNoise=high
   RC1_GeneralVal=very_high 42 conf:(0.95)
3. T1D_VisualNoise=normal T2D_PullDownMenus=high T3D_LinkStyle=high 34
   ==> RC1_GeneralVal=very_high 36 conf:(0.94)
4. T2D_LinkStyle=high T4D_PageStyle=high 34 ==> GeneralVal=very_high 37 conf:(0.91)
5. T1D_PageStyle=low T1D_Animation=very_high ==> T1D_HomogeneousStyle=high
   RC1_GeneralVal=high 22 conf:(0.90)
```

Fig. 2. Ranked list R#1 corresponding to Cognitive Walkthrough #1 and the Design Category. Third Stage of *UsabAIPO Project* (visualization provided by WEKA).

Usab. Prob Pattern #1	Lack of an homogeneous style in different pages
Frequency (1 to 5)	4
Justification	AR#1, AR#4 and AR#5 in ranked list R#1
Ev. Team Comment	Usability problem pattern observed in 35 webs
Recommendations	Include a study of the corporative image of the university and use results as guidelines for the webpage graphic design.

Fig. 3. Report of usability problem pattern. Third Stage of *UsabAIPO Project*.

Similar association rules were grouped and jointly rephrased. The frequency of every pattern was calculated on the basis of a weighted percentage which represents the number of rules supporting the pattern adjusted in function of the support and confidence of the rule. The explanation was stated by a list of alphanumeric characters identifying *ARs* and ranked lists supporting the usability problem pattern, as well as the corresponding comment of the evaluation team. An internal report describing the results was also written [8].

5 Conclusions. Related and Future Work

Identifying usability problem patterns in a given context of use is a challenging problem. However, *QUT* methods for qualitative usability evaluation are focused on considering particular systems, becoming inappropriate when analyzing a given context of use C as a whole on the basis of the joint analysis of a large number of systems in C. To cope with this situation, this paper presents the $QUTC_{KDD}$ methodology, which is based on the integration of the traditional *QUT* methods and *KDD* techniques. Note that in $QUTC_{KDD}$ intuitions that were usually informally stated by the evaluation team during the *QUT* process can now be appropriately analyzed in the context of *KDD*-based techniques. Note also that the evaluation can be enriched with the detection of hidden relationships among qualitative data that are documented with a formal basis.

To the best of our knowledge, there are no other approaches to extend *QUT* by incorporating *KDD* techniques as presented here. Typical models for usability evaluation under *UE* are not focused on the evaluation of contexts of use as a whole [2,4,7,11,13,14,18,20]. However, there exist some approaches which include

AR mining for assessing usability oriented towards the evaluation of a single system. For example, the AWUSA framework [19] presents an automatic tool for evaluating the usability of a web site by combining logging techniques and datamining, along with the static structure of the web site. Another approach where the *ARs* are used to test the usability of a single system is described in [1]. In this case, logging techniques are based on browsing activities performed by users. Techniques for statistical analysis (such as correlation) can also be applied to find relationships in usability testing on the basis of the available data. However, note that correlation refers to the departure of a group of variables from independence, but does not imply causality, as in the case of *ARs*. In addition, note that in our proposal the variables to be analyzed do not need to be identified explicitly, as they are found automatically by the *AR* mining algorithm.

Part of our future work is focused on testing different ranking functions for *ARs*, as suggested in [3]. To achieve this, the use of more powerful *KDD* platforms (e.g., KDDML platform [17]) should be crucial, as this kind of platform can be seen as an evolution of *KDD* engines like WEKA [17]. Another research line currently explored is the development of a usability evaluation module based on $QUTC_{KDD}$ capable to be integrated in a *Usability Evaluation Management System*. Particularly, an integration with SketchiXML [5] is under consideration.

Acknowledgments. This work was supported by Projects TIN2004-08000-C03-03 (FEDER, European Union) and SGR-00881 (DURSI, Catalonian Government, Spain).

References

1. J. Alipio, J. Poças, and P. Azevedo. Recommendation with Association Rules: a web mining application. *Data Mining and Warehouses Conf. IS-2002*, 2002.
2. R. Bailey, R. Molich, J.Dumas, and J.M. Spool. Usability in Practice: Formative Usability Evaluations. *ACM CHI2002*, 2002.
3. D. Choir, B. Ahn, and S. Kim. Priorization of association rules in DM: Multiple criteria decission approach. *Expert System with Applications*, 29:867–878, 2005.
4. L. Constantine and L. Lockwood. *Software for Use. A practical Guide to the Models and Methods of Usage-Centered Design.* Addison-Wesley, 1999.
5. A. Coyette and J. Vanderdonckt. A Sketching Tool for Designing Anyuser, Anyplatform, Anywhere User Interfaces. In *Lecture Notes in Computer Science. Proc INTERACT 2005*, volume 3585, pages 550–564. Springer Verlag, 2005.
6. J. Demsar, B. Zupan, and G. Leban. Orange: From experimental machine learning to interactive data mining. Technical report, University of Ljubljana, 2004.
7. J. S. Dumas and J. C. Redish. *A Practical Guide to Usability Testing.* Intl. Specialized Book Service Inc, 2000.
8. M. P. González, J. Lorés, R. Gabandé, and T. Granollers. UsabAIPO Project. 3rd Stage. Technical report, AIPO Association, 2006.
9. M. P. González, J. Lorés, A. Pascual, and T. Granollers. Evaluación Heurística de Sitios Web Académicos Latinoamericanos dentro de la Iniciativa UsabAIPO. In *Proc. del VII Int. Interacción Persona-Ordenador*, 2006.
10. J. Han and M. Kamber. *Datamining: Concepts & Techniques.* M. Kaufmann, 2000.

11. Andreas Holzinger. Usability engineering for software developers. *Communications of the ACM*, 48(1):71–74, 2005.
12. T. Imielinski and A. Virmani. MSQL: A Query Language for Database Mining. *Data Min. Knowl. Discov.*, 3(4):373–408, 1999.
13. M. Y. Ivory and M. A Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.*, 33:470–516, 2001.
14. D. J. Mayhew. *The Usability Engineering Lifecicle. A practioner's handbook for user interface desing.* M. Kaufmann, 1999.
15. T. Mitchell. *Machine Learning.* McGraw Hill, 1997.
16. D. Pyle. *Data Preparation for Data Mining.* M. Kaufmann, 1999.
17. A. Romei, S. Ruggieri, and F. Turini. KDDML: a middleware language and system for knowledge discovery in DB. *Data and Knowledge Eng.*, 57(2):179–220, 2006.
18. J. Scholtz. Adaptation of traditional usability testing methods for remote testing. In *HICSS '01: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 5*, page 5030. IEEE Computer Society, 2001.
19. T. Tiedtke, C. Märtin, and N. Gerth. Awusa. A tool for automated website usability anlaysis. *9th Int. Workshop DSVIS*, 2002.
20. E. van Veenendaal. Low-cost usability testing. *Software Quality and Software Testing in Internet Times*, pages 153–164, 2002.
21. C. Wharton, J. Rieman, C. Lewis, and P. Polson. *Usability Inspection Methods*, chapter The Cognitive Walkthrough Method: A practitioners guide, pages 105–140. John Wiley & Sons Inc, 1994.
22. Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques.* M. Kaufmann, 2005.

Usability Constructs: A Cross-Cultural Study of How Users and Developers Experience Their Use of Information Systems

Morten Hertzum[1], Torkil Clemmensen[2], Kasper Hornbæk[3], Jyoti Kumar[4], Qingxin Shi[2], and Pradeep Yammiyavar[4]

[1] Computer Science, Roskilde University, Roskilde, Denmark
mhz@ruc.dk
[2] Department of Informatics, CBS, Copenhagen, Denmark
{tc.inf,qs.inf}@cbs.dk
[3] Department of Computer Science, University of Copenhagen, Denmark
kash@diku.dk
[4] Indian Institute of Technology, Guwahati, India
{jyoti.k,pradeep}@iitg.ernet.in

Abstract. Whereas research on usability predominantly employs universal definitions of the aspects that comprise usability, people experience their use of information systems through personal constructs. Based on 48 repertory-grid interviews, this study investigates how such personal constructs are affected by two factors crucial to the international development and uptake of information systems: cultural background (Chinese, Danish, or Indian) and stakeholder group (developer or user). We find that for the user group frustrating and useful systems are experienced similarly, whereas for the developers frustrating systems are experienced similarly to easy-to-use systems. Looking at the most characteristic construct for each participant we find that Chinese participants use constructs related to security, task types, training, and system issues, whereas Danish and to some extent Indian participants make more use of constructs traditionally associated with usability (e.g., easy-to-use, intuitive, and liked). Further analysis of the data is ongoing.

Keywords: Cultural usability, Usage experiences, Repertory-grid technique.

1 Introduction

The concept of usability has been debated for decades [10, 16]. Most of this work, however, defines usability analytically or by reference to standards such as ISO 9241-11 [12]. Conversely, we know little about how people talk about their experiences with the systems they commonly use. Following Kelly [13] we take descriptions of such use experiences as indicative of the personal constructs people employ in relating to systems. By recognizing the personal nature of such usability constructs we seek to avoid unwarranted universalism and to explore how usability constructs are affected by two factors crucial to the international development and uptake of systems:

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 317–326, 2007.
© Springer-Verlag Berlin Heidelberg 2007

- *Cultural background*. The first aim of this study is to contribute to an elaboration of the cultural aspects of usability by investigating whether similarities and differences in people's usability constructs correlate with their cultural background. Cultural background is, in this study, taken to mean people's country of birth and residence. Though cultural usability is emerging as a topic [2, 9, 14], culture has typically not been considered at all in commonly accepted usability definitions.
- *Stakeholder groups*. The second aim of this study is to compare and contrast users' and developers' usability constructs. Any systematic differences in the usability constructs of these two stakeholder groups might impede user-developer communications about user requirements or system evaluations. Additionally, systematic differences may serve to elaborate and bridge between existing usability definitions.

To investigate the two factors empirically, we conduct repertory-grid interviews with users and developers with three different cultural backgrounds (Chinese, Danish, and Indian) and analyse the data descriptively and by means of principal-component analysis.

2 Related Work

While Barber and Badre [2] argue that users' cultural background can directly impact their performance using information technology (IT), the nature of this merging of culture and usability is presently far from clear. Research provides evidence that users' beliefs about their acceptance of systems and users' actual use of systems may be influenced by their cultural background. For example, Evers and Day [4] found that Chinese students attached more importance to perceived usefulness in forming an opinion about whether to accept a system, compared to Indonesian students who attached more importance to perceived ease of use. Honold [9] showed that washing machines were used quite differently in Indian and German households and that these differences led to fundamentally different user requirements. A prominent attempt at explaining the dimensions along which cultures differ is Hofstede's work [7], which identified five cultural dimensions: power distance, uncertainty avoidance, individualism/collectivism, masculinity/femininity, and long-term/short-term orientation. Hofstede's work has, for example, been introduced in HCI by Marcus and Gould [14] in relation to web-site design.

With respect to stakeholder groups, it is well-recognized that users and developers differ in manifold ways but, to our knowledge, no studies have systematically compared and contrasted how users and developers understand usability. Other stakeholder groups' understanding of usability have, however, been compared. Morris and Dillon [15] found that usability was not a central concern to managers responsible for making decisions about which IT systems to procure, but that it was a central concern for the end users. Moreover, managers and end users tended to conceptualize usability in different ways. To the managers, usability was predominantly a feature of the IT systems, such as 'having a point-and-click interface'. To the end users, usability was also dependent on the interactions among users, tasks, tools, and context. For example, one end user defined usability as "being able to use the software to perform the

tasks needed without excessive consultation" [15: p 253]. Holcomb and Tharp [8] had users rank the importance of the individual elements in a model of usability. Functionality was rated significantly more important than the six other elements of the model, namely consistency, user help, naturalness, user control, feedback, and minimal memorization. As the users had no option for extending the model with additional elements it was, however, not possible to say whether the model captured what the users considered to be the important elements of usability.

The repertory-grid technique, which is used in this study, originates from Kelly's personal-construct theory [13]. He rejected the idea that people perceive and make sense of their world by means of conceptions that exist independently of the individual person and instead proposed that people see their world through a set of personal constructs. These personal constructs are created over time in the course of people's interactions with their environment and express the dimensions along which a person can differentiate among objects and events. That is, each construct is defined by a similarity-difference dimension. Kelly [13] devised the repertory-grid technique to elicit personal constructs in the context of psychological counselling. The technique has subsequently been used successfully in interviews aiming to capture users' thoughts about IT products [1, 6, 17] and suggested for use in cross-cultural studies of information systems [11].

3 Method

To investigate the constructs people use to describe their experience of the information systems they use, we conducted repertory-grid interviews with people from two stakeholder groups (developers and users) and with three cultural backgrounds (Chinese, Danish, and Indian).

3.1 Participants

For each combination of stakeholder group and cultural background, we interviewed eight people. The Chinese participants lived and were interviewed in Beijing, the Danish participants in Copenhagen, and the Indian participants in Bangalore, Guwahati, Hyderabad, and Mumbai. Table 1 summarizes the 48 participants' gender, age, and IT experience. The participants had average to excellent English skills.

Table 1. Participant profiles

Group	*Gender*		*Age (years)*		*IT experience (years)*	
	Male	*Female*	*Mean*	*SD*	*Mean*	*SD*
Chinese developers	5	3	31.5	1.9	10.6	1.7
Chinese users	5	3	27.3	1.9	8.4	1.9
Danish developers	5	3	36.6	5.8	19.3	5.8
Danish users	5	3	36.8	6.2	16.9	3.6
Indian developers	8	0	29.6	1.7	9.9	2.5
Indian users	5	3	29.0	4.0	7.0	2.1

3.2 Procedure

Participants were interviewed individually by a person with a cultural background similar to their own. First, the study was described to the participant and the repertory-grid technique explained. Second, participants filled out a questionnaire about their background and signed an informed-consent form. Then, participants tried to elicit constructs with the repertory-grid technique on a couple of training tasks. After these preparatory steps, the actual repertory-grid interviews were conducted. They consisted of two steps: selection of systems and elicitation of constructs.

In selecting systems, the participant was asked to consider "*the array of computer applications you use for creating, obtaining, revising, managing, and communicating information and documents in the course of your day-to-day activities*." This included applications the participants use regularly but excluded applications they had only used once or twice and applications they merely know of. On this background participants were asked to select a system within each of six categories: my text processing system, my email, a useful system, an easy-to-use system, a fun system, and a frustrating system.

In eliciting constructs, the participant was successively presented with groups of three of the selected systems and asked: "*Can you think of some important way in which your personal experience using these three systems makes two of the systems alike and different from the third system?*" Having indicated the two similar systems, the participant wrote down a short phrase that told how these two systems were alike – the construct – and another short phrase that told how the third system differed – the contrast. Then, a seven-point rating scale was defined with this construct-contrast pair as its end points, and the participant rated all six systems according to this rating scale.

This procedure was repeated for all twenty combinations of three systems, in random order, or until the participant was unable to come up with a new construct for two successive combinations. The interviews were conducted in the participants' native language, if participants preferred that, or in English. Constructs and their contrasts were always recorded in English. In accordance with cultural customs, Danish and Indian participants received no compensation for their participation in the study while Chinese developers were paid 200RMB for their participation and Chinese users 50RMB. Each interview lasted about 1.5 hours.

3.3 Interviewer Preparations

The repertory-grid interviews were conducted by three of the authors. Three activities were performed to ensure that they conducted their interviews in the same way: First, we wrote an interview manual with step-by-step instructions about how to conduct the interviews. Second, each interviewer conducted a pilot interview. Third, we met before the pilot interviews to walk through a draft version of the interview manual and again after the pilot interviews to discuss experiences gained from the pilot interviews. The outcome of these preparations was the final version of the interview manual and a common understanding among the interviewers about how to conduct the interviews.

4 Results

We first present the participants' choice of systems and analyse the constructs used by individual participants. Next we analyse differences among systems, between stakeholder groups, and across participants' cultural backgrounds.

4.1 Participants' Choice of Systems

The 48 participants each selected six systems to be used in the elicitation of constructs. In the category 'my text processing system', 44 participants selected Microsoft Word; the remaining participants were divided on four additional systems. In the category 'my email', 20 participants selected Microsoft Outlook and seven participants selected Yahoo; the remaining participants were divided on seven additional systems. For the four other categories the participants selected a more mixed variety of systems. In the category 'a useful system' the most frequently selected system was Google (5 participants) and 36 additional systems were selected by one to four participants. In the category 'an easy-to-use system' Internet Explorer (5 participants) was the most frequent of a total of 30 different systems. In the category 'a fun system' three systems were selected by three participants (Google, Powerpoint, and Yahoo Messenger) and 32 additional systems by one or two participants. Finally, in the category 'a frustrating system' the most frequently selected system was Microsoft Excel (3 participants) and 42 additional systems were selected by one or two participants.

4.2 Constructs Used by Individual Participants

Participants reported an average of 13.8 constructs (SD = 3.6). The constructs varied much across individual participants in their level of abstraction, reference to personal experience, and relation to specific applications. Table 2 shows a summary of the most characteristic constructs as identified by principal-component analyses of individual grids. For each such analysis we selected the construct corresponding to the component that explained the largest amount of variance [5: pp 86-87, 3: p 14], for a total of 48 constructs.

Table 2. Participants' most characteristic construct. The table shows the most characteristic constructs that are shared by three or more of the 48 participants.

Most characteristic construct	*No. of participants*
Easy-to-use vs. Difficult	5
Work vs. Fun	5
Need for training vs. Walk-up-and-use	3
For myself vs. For the public	3
Simple vs. Complex	3

Across all 661 constructs, prominent kinds of construct relate to performance (e.g., 'Fast'), security (e.g., 'Easy to be affected by virus'), social issues (e.g., 'Communicate with other people'), frequency of use (e.g., 'Use everyday'), the context of use (e.g., 'Can use away from my desk'), the need to update and install programs (e.g., 'No need to update'), hedonic quality (e.g., 'Happy', 'Lot of fun to use'), aesthetics (e.g., 'Colourful interface'), and forgivingness (e.g., 'Insensitive to small mistakes').

4.3 Differences Among System Types

Fig. 1 shows the result of an individual differences multi-dimensional scaling on the six system types, across all 48 grids [5: p 99]. System types appear close together on the figure if participants rated them similarly on the rating scales defined by the construct-contrast pairs and far apart if participants rated them differently. The most noteworthy observation from this analysis is that the useful system and the frustrating system are close together, suggesting that participants rated these systems similarly. This observation is confirmed by an analysis of correlations of ratings among systems showing that ratings of frustrating systems are negatively correlated with ratings of all system types (r = -.14 to -.31, all ps < .001), except the useful system (r = .028, p > .4). This is not to say that frustrating systems are useful, but merely that usefulness does not indicate absence of frustration. For 25% of the 661 constructs, the ratings of the frustrating and the useful system are identical.

Fig. 1 also indicates that participants rate easy-to-use and fun systems similarly. Along one of the two dimensions in the multi-dimensional scaling easy-to-use and fun systems are also rated in opposition to useful and frustrating systems.

4.4 Differences Between Stakeholder Groups

Fig. 2 suggests that the two stakeholder groups conceptualize the systems differently. One difference is that for developers the frustrating system is close to the easy-to-use

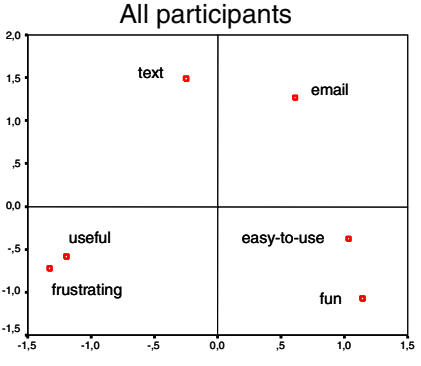


Fig. 1. Multi-dimensional scaling of system types based on data from all 48 participants. The stress value – an indicator of how well the scaling fits the raw data – for this scaling is .379.

system; this association is not found for the user group. Correlations of raw ratings show that easy-to-use and frustrating systems are not significantly correlated for developers (r = -.11, p > .05), but have a significant negative correlation for users (r = -.22, p < .001). For the user group we find a relation between the frustrating system and the useful system similar to that discussed in Section 4.3. An explanation of the difference between stakeholder groups for easy-to-use and frustrating systems may be that easy-to-use systems often cannot match the complexity of developers' work tasks and therefore resemble systems that cause developers frustrations. Another explanation may be that developers have higher or different standards for what constitutes an easy-to-use system. These explanations are merely tentative for three reasons: the dimensions of the plots in Fig. 2 are not easily comparable, the systems chosen as frustrating vary considerably across participants, and the constructs used by the two stakeholder groups may differ.

Another difference between the two stakeholder groups is that email seems to resemble text-processing systems for the developers, whereas for the user group email shares many of the properties of easy-to-use systems.

4.5 Differences Across Cultural Backgrounds

Fig. 3 shows a separate multi-dimensional scaling for participants with each of the three cultural backgrounds. From the diagrams it seems that systems for text and email are construed differently across cultures. In contrast to Danish and Indian participants, Chinese participants seem not to associate text-processing and email systems with each other (r = .008, p > .8), possibly reflecting a different role of email or issues associated with the support for writing Chinese. A further difference is that the fun system is associated with a different system for each of the participants' cultural backgrounds: for Chinese participants it is email, for Danish participants it is the easy-to-use system, and for Indian participants it is the useful system.

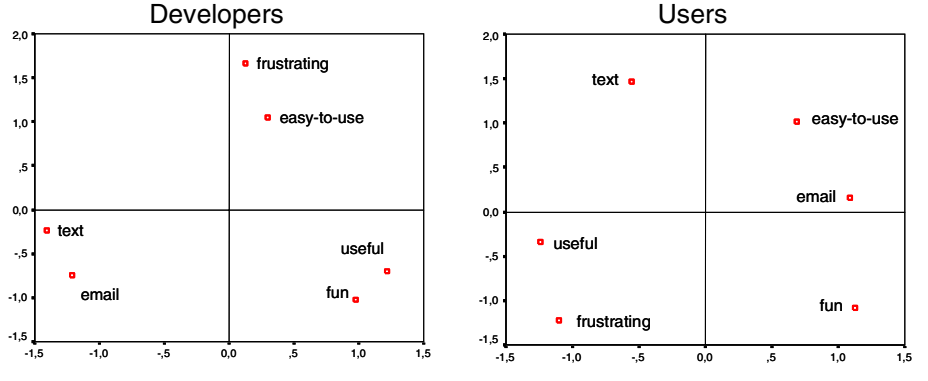


Fig. 2. Multi-dimensional scaling of system types: left panel is based on data from the 24 developers, and right panel is based on data from the 24 users. The stress values for these scalings are .385 and .311.

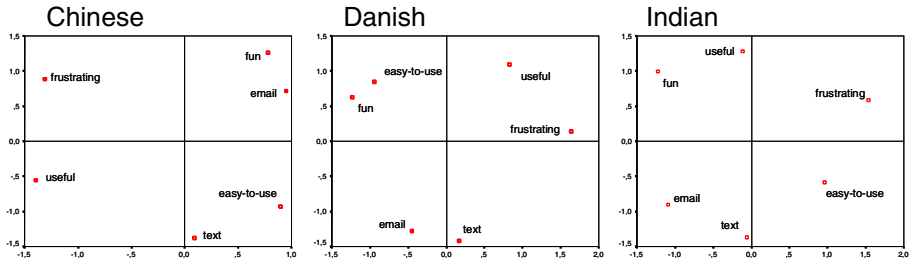


Fig. 3. Multi-dimensional scaling on participants' cultural background: left panel is based on data from the 16 Chinese participants, middle panel is based on data from the 16 Danish participants, and right panel is based on data from the 16 Indian participants. The stress values for these scalings are .319, .331, and .331.

Table 3 suggests that participants' cultural background influences which constructs they employ. Chinese participants have as their most characteristic construct a range of issues related to security, task types, training, and system issues. In contrast, Danish and to some extent Indian participants seem to mention more frequently aspects traditionally associated with usability (e.g., 'Easy-to-use', 'Intuitive', and 'Liked'). Eight (Danish) and six (Indian) of the most characteristic constructs are of this kind, as opposed to none of the constructs elicited by Chinese participants. Further, a distinction between work and leisure activities is more widely reported by Indian participants. Among all 661 constructs, however, the number of constructs that can be related unequivocally to this distinction are 15 (Indian), 11 (Chinese), and 12 (Danish).

Table 3. The most characteristic construct for each participant, divided onto cultural background. Some constructs have been slightly rephrased to be intelligible out of context.

Chinese participants	*Danish participants*	*Indian participants*
Often bring virus to computer	Experienced	Creative
Are used mostly by professionals	Stable and robust	Straight-forward
Daily use	Stand-alone program	Helps structuring
Used for email	Supports browsing	Natural way of use
Automatic installation	Give overview	Intuitively trustworthy
Use for programming	Context help	Complex product
Infrequent updating	Single supplier of application	Simpler
Use it when chatting	Simple	Stand alone application
Need internet connection	Easy-to-use	Just for relaxing
Can input information	Support numbers and figures	Help available
Can create something with applications	Easy-to-use	Entertainment
Need to use id	Intuitive	For work
Need training	Give focus	Recreation
Have many users	Process information	Liked
Need more memory	More complicated	Effective tools
Can use it first time	Creative	Related to public

5 Discussion and Conclusion

The participants in this study made use of a rich variety of constructs in talking about their experiences using IT systems. Following Kelly [13] these constructs, and their associated contrasts, define the dimensions along which participants perceive and are able to differentiate among usage experiences with different systems. Hence, the constructs can be seen as the building blocks of the participants' personal concepts of usability. In this sense the constructs stand in contrast to most definitions of usability, in which usability is defined analytically or with reference to standards like ISO 9241-11 [12]. An implicit assumption of these definitions is that they are valid across stakeholder groups and persons with different cultural backgrounds. Our analysis suggests that this assumption may not hold.

In this study, 48 participants made use of 661 construct-contrast pairs in describing how their experiences using some systems are alike and different from their experiences using other systems. Some of the constructs used by participants fit well with common definitions of usability, for example by emphasizing ease-of-use. Other constructs are well-known to human-computer interaction in that they describe use situations, the need for training, or frequency of use. However, a number of the elicited constructs are hard to reconcile with prevailing definitions of usability. For example, participants frequently mentioned issues of security – relating both to viruses and trustworthiness. The distinction between work and leisure is another example of a construct frequently employed by participants in distinguishing among systems but mostly glossed over in models of usability [e.g., 10, 12].

Some of the differences in the constructs employed by participants appear to be related to participants' cultural background and stakeholder group. A fun system is experienced similarly to email by Chinese participants, similarly to easy-to-use systems by Danish participants, and similarly to useful systems by Indian participants. The most characteristic construct for each participant provides further evidence for cultural differences in how the use of IT systems is experienced. Whereas traditional usability aspects, such as intuitiveness, are frequent among the most characteristic constructs of Danish and to some extent Indian participants, they are absent for the Chinese participants. This suggests cultural variation in the participants' concept of usability. In addition, developers seem to experience frustrating systems similarly to easy-to-use systems, whereas users experience frustrating systems similarly to useful systems. This adds to previous work by Morris and Dillon [15] and points toward possible sources of confusion in user-developer communication.

The present study has a number of limitations. First, the repertory-grid interviews were conducted by three interviewers. This may have introduced subtle differences in how interviews were conducted though we tried to avoid this through careful interviewer preparations. We chose against having the same interviewer for all interviews because it would mean that most or all participants would be interviewed by a person with a cultural background and native language different from their own. Second, some of the elicited constructs cannot readily be interpreted as aspects of the participants' experiences using the systems (e.g., 'Can have many windows'). However, in the absence of clear criteria for when to exclude a construct we included all constructs in the analysis. Third, part of our analysis is based on the most characteristic construct for each participant and, thus, disregards all additional constructs elicited by the participants.

Further analysis, including content analysis of the constructs, is ongoing.

Acknowledgements. This study was co-funded by the Danish Council for Independent Research through its support of the Cultural Usability project. We are grateful to the interviewees, who participated in the study in spite of their busy schedules.

References

1. Baber, C.: Repertory grid theory and its application to product evaluation. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 157–165. Taylor & Francis, London (1996)
2. Barber, W., Badre, A.: Culturability: The merging of culture and usability. In: Proceedings of the Fourth Conference on Human Factors and the Web. AT&T Labs, Basking Ridge, NJ (1998). Available at: http://zing.ncsl.nist.gov/hfweb/att4/proceedings/barber/index.html
3. Bell, R.C.: Using SPSS to Analyse Repertory Grid Data. University of Melbourne, Melbourne, AU (1997). Available at: http://eu.wiley.com/legacy/wileychi/fransella/supp/gridspssman.doc
4. Evers, V., Day, D.: The role of culture in interface acceptance. In: Howard, S., Hammond, J., Lindgaard, G. (eds.) INTERACT '97: Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction, pp. 260–267. Chapman and Hall, London (1997)
5. Fransella, F., Bell, R., Bannister, D.: A Manual for Repertory Grid Technique, 2nd edn. Wiley, New York (2004)
6. Hassenzahl, M., Wessler, R.: Capturing design space from a user perspective: The repertory grid technique revisited. International Journal of Human-Computer Interaction 12(3&4), 441–459 (2000)
7. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage, Thousand Oaks, CA (2001)
8. Holcomb, R., Tharp, A.L.: What users say about software usability. International Journal of Human-Computer Interaction 3(1), 49–78 (1991)
9. Honold, P.: Culture and context. An empirical study for the development of a framework for the elicitation of cultural influence in product usage. International Journal of Human-Computer Interaction 12(3&4), 327–345 (2000)
10. Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human-Computer Studies 64(2), 79–102 (2006)
11. Hunter, M.G., Beck, J.E.: Using repertory grids to conduct cross-cultural information systems research. Information Systems Research 11(1), 93–101 (2000)
12. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. ISO, Genève, CH (1998)
13. Kelly, G.A.: The Psychology of Personal Constructs. Norton, New York (1955)
14. Marcus, A., Gould, E.W.: Crosscurrents: Cultural dimensions and global web user-interface design. ACM Interactions 7(4), 32–46 (2000)
15. Morris, M.G., Dillon, A.P.: The importance of usability in the establishment of organizational software standards for end user computing. International Journal of Human-Computer Studies 45(2), 243–258 (1996)
16. Shackel, B.: The concept of usability. In: Bennett, J., Case, D., Sandelin, J., Smith, M. (eds.) Visual Display Terminals: Usability Issues and Health Concerns, pp. 45–87. Prentice-Hall, Englewood Cliffs, NJ (1984)
17. Tan, F.B., Hunter, M.G.: The repertory grid technique: A method for the study of cognition in information systems. MIS Quarterly 26(1), 39–57 (2002)

A Study for Usability Risk Level in Physical User Interface of Mobile Phone

Beomsuk Jin, Sangmin Ko, Jaeseung Mun, and Yong Gu Ji*

Yonsei University., 134 Sinchon-Dong, Seodaemun-gu, Seoul, Korea
{kbf2514jin,sangminko,mjs,yongguji}@yonsei.ac.kr

Abstract. The purpose of this study is to develop a framework of quantitative evaluation of PUI risk level to ensure the usability in designing mobile devices. Three PUI factors—key type, use scene and device form—were selected as the main criteria for PUI risk level. They are defined as Key Manipulation Value (KMV), Function Manipulation Value (FMV) and Handling Value (HV), considering the requirements. In short, this study provides a framework of quantitative evaluation with the requirements of the three PUI factors, and analyzes risk level by KMV, FMV and HV. This result can be utilized as a criterion for usability at the design phase. In addition, evaluation with this framework at the early design phase helps to anticipate the problems, so the opportunity to solve the problem can be offered in advance.

Keywords: Mobile Phone, Physical User Interface, Risk Level.

1 Introduction

These days, many electronic products are rapidly improved by development of digital and telecommunication technologies. As various functions are added in digital products, product's UI, exterior design, application and usability are affected by these additional changes. Especially, mobile devices are no longer only a device for calling and sending SMS. New functions such as camera, game, DMB, GPS and wireless internet have been established as core components by development of advanced software and hardware technology. This multimedia device has become feasible through digital convergence. Also it became a core-device to satisfy users' various requirements in many fields such as entertainment, business and information [4]. However, as various functions are added into one small device, complexity of the device is causing inefficiency in device control [7]. Therefore UI design that can support user satisfaction and ease of use is getting important.

In general, UI is classified with Graphic User Interface (GUI) and Physical User Interface (PUI). PUI is the term that includes practical and physical characteristics which is related to device's exteriors like buttons, switches and levers to manipulate the device. This must be concerned in early design process and is highly related to context of GUI, which executes applications through display and gives a feedback

* Corresponding author.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 327–335, 2007.
© Springer-Verlag Berlin Heidelberg 2007

from its execution [2]. Desktop users devote all of their visual resources to the application which they are interacting. In contrast users of mobile devices, are typically in motion while they are using their device, can not devote their all resources to interact with mobile application. Moreover, the mobile devices are getting smaller and more multi-functional so the form types are expanded from basic form type, bar type, folder type and slide type to advanced and mixed type, swing and swivel type. This change causes difficulties in use due to limited screen real estate and limitation in design of physical buttons [1]. Thus, multi-functional and minimized mobile device has more problems than other digital devices, which leaves an important challenge to overcome the limitation of efficient control for using many functions of devices. Mobile device is a unique type of digital convergence appliance in that PUI and GUI are combined. Therefore, it must be designed with advanced paradigm as preceding ergonomics about hand-tools [6].

In this paper, ease of use is evaluated in the aspect of PUI of calling, SMS, camera, MP3 and DMB using mobile device that is affected by mobile device's physical components(device type, button type and button position). Three PUI evaluation factors are selected: key-type, use-scene, form-factor. Key-type evaluates the degree of efficiency about performing task, use-scene evaluates controllability due to button region and button type, and form-factor evaluates the degree of interruption among key types. The three PUI evaluation values are defined as Key Manipulation Value (KMV), Function Manipulation Value (FMV), Handling Value (HV), respectively. Each of three evaluation values generated mobile device's PUI risk level by estimating requirements related to PUI. Also, the alternatives, which can solve the problem of PUI, are made by analyzing the reasons of high risk level. As a result, mobile device's risk level can be evaluated by three defined PUI evaluation values - KMV, FMV and HV. By these, PUI evaluation framework is generated, which can find the predictable problems of ease of use and manage the problems in early design step.

2 Literature Review

Previous researches were examined to extract requirements about key-type, use-scene, and form-factor.

In most of previous studies about requirements of Key-type, improving user performance in each task was the main issue. Nielsen studied supporting visual and tactile feedback to improve tool's controllability[5]. Also, the necessity of error prevention which reduces errors occurred from user's mistakes is mentioned. These are main requirements for hand tool device. Shneiderman researched to improve menu navigation for structural and nonstructural information searching by measuring frequency of button usage, accessibility for manipulation, user satisfaction [12]. This research is about most efficient button which can support user in discrete and continuous tasks. And the result can be applied to the navigation button for menu navigation. Furthermore, the Kasper divided difference of multidimensional control and unidimensional control, and different applicable range of various control types related to discrete control or continuous control were analyzed in this paper [9].

About interface design of mobile device's keypad, efficient text input and control method were studied, and intuitive and efficient guideline for key arrangement was researched. A new advanced input method was proposed with consideration of balancing input efficiency, ergonomics, usability and cost [11]. Also text input using mobile device is emphasized and a research about comparison and evaluation by calculated input time by Fitt's law was done by Silverberg [13]. Even though it needs little force to manipulate device, user can feel fatigue by finger angle in short time. For that reason, maximum muscular strength which is changed by finger angle during key control with grasping device is also researched [3]. In this research, interface design, which is considered button controllability, accuracy and interference, is required.

Lastly, there were many researches about finger's muscular strength that is related to hand grip and button control design of electronic product in ergonomics' view. This paper focused not only on tool grip design but also on optimizing the best knob shape and size, grip force and grip type using anatomic structure of hand [8]. Miniaturization of mobile device will bring difficulty at the input method. For that reason, Nambu Hirotaka proposed that users need to grip different part of device when using the right bottom part of mobile device [10]. When user tries to input character continuously, the grip-stability with some friction can provide comfort to users; additional research about a bottom part of mobile device's grip is performed.

3 Risk Level Evaluation Methodology

Evaluation of mobile device's risk level consists of 4 steps. In the first step, mobile device's features of existing mobile phones are analyzed. The form of device and motion of each form are analyzed as well. Also types of each key and main keys for controlling calling, SMS, camera. MP3 and DMB were investigated. In second step, 3 PUI evaluation factors(key-type, use-scene and form-factor) requirement are collected through previous researches and literature research to estimate the requirement's weight. In third step, the values of KMV, FMV and HV are defined to calculate the risk level quantitatively. Finally, the last step evaluated PUI risk level of mobile device according to evaluation framework of risk level (Fig. 1).

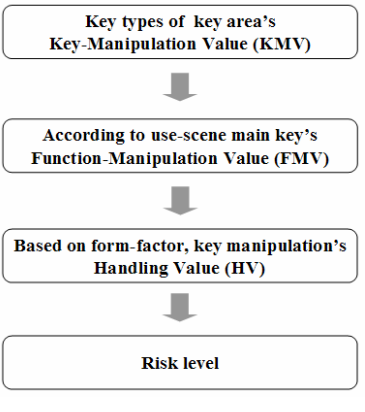


Fig. 1. Framework of PUI risk level

3.1 Feature Analysis of Mobile Devices

Physical movements of different device types are analyzed for evaluating PUI risk level. 133 products (Domestic product: 101, Foreign product: 32) were analyzed. Mobile devices are divided into 4 kinds of form-type(bar, folder, slider and swing). And each different form-types of devices had 13 physical transformations (Table 1). Bar-type consist of normal bar-type and mixed bar-type. Normal bar-type is bar-type without transformation. And mixed bar-types are combined with swing or swivel type. Folder-type consists of normal type (up type) and abnormal type (left-right, up-swing and up-swivel type). Slide-type is divided into up, down and up-down type. And there are mixed slide-types which are combined with up type and swing type or swivel type. Normal swing-type is also founded. We defined two positions of mobile devices. The position without transformation is defined as 'base position' and the position with ion is defined as 'home position'.

Table 1. Form of mobile devices

Form-type	Form-factor(Movement)
Bar	Bar
	Bar + Swing
	Bar + Swivel
Folder	Up
	Right and Left
	Up + Swing
	Up + Swivel
Slider	Up
	Up + Down
	Down
	Up + Swing
	Up + Swivel
Swing	

Mobile device is divided into navigation area, function area, numeric area and side area, and key-type of each area were investigated (Table 2). And dorm key, touch key, jog-disk, jog-stick and wheel are used in each key-area of mobile device.

Table 2. Key-type in each key-area

Key-area	Key-type
Navigation	Dorm key, Touch key, Jog-disk, Jog-stick, Wheel
Function	Dorm key, Touch key, Jog-disk, Jog-stick
Numeric	Dorm key, Touch key, Wheel
Side	Dorm key, Touch key, Jog-disk, Jog-stick, Wheel

Generally in mobile device, there are various functions. However in this research 5 main functions(calling, SMS, camera, MP3 and DMB) which have high usage

frequency were selected to be evaluated. Also main-key in each key-area was selected to analyze the key-area which is needed to control each functions (Table 3).

Table 3. Main-key and key-area in each function

Function	Main-key	Key-area
Calling	Numeric key	Numeric
	Calling, clear, end key	Function
	Volume control key	Side
Short Message Service (SMS)	Numeric key	Numeric
	Mode switch key	Function
	Clear, confirmation key	Function
	Specific letter key	Navigation
Camera	Shutter key	Function
		Side
	Zoom key	Navigation
		Side
	Brightness control key	Navigation
		Side
MP3	Play, Stop key	Function
	Music search key	Navigation
	Volume control key	Navigation
		Side
DMB	Channel switch key	Navigation
		Numeric
	Volume control key	Navigation
		Side

As a result of investigation and analysis, classified form-type of mobile device, key-type in each key-area and main-keys in each function were selected as evaluation components.

3.2 Requirement Collecting and Weight Assessment

Requirements asked in PUI factors (review, key-type, use-scene and form-factor) are collected through literature review. And by selection process, 16 requirements are selected; 8 of key-type, 6 of use-scene and 2 of form-factor (Table 4).

Requirement's weight of Key-type was estimated by considering main goal and task of each key-area. About use-scene, requirement's weight was estimated by considering key controllability, performance of key control and interference in key control. Also requirement's weight of Form-factor was estimated by considering interference and stability between form of mobile device and motion of key-type.

The weights of each requirement were verified by HCI experts and mobile device designers' discussion.

Table 4. Requirement and definition

PUI factors	Requirement		Definition
Key-type	Feedback		Provide tactile feedback using control keys or not
	Quick navigation		Degree of providing shortcut to navigate in menu which is consist of many list
	Detail control		Degree of providing detail control of small numeric unit (ex: volume, zoom)
	Eye-tracking		Degree of providing key manipulating without eye-tracking
	Multidimensional control		Degree of providing multidimensional control in 2 levels
	Error		Degree of providing accurate key manipulation
	Thumb range		Degree of providing natural key manipulation in Thumb range
	Task performance		Degree of performing task
Use-scene	Thumb range according to main used key	Stability	Degree of grip-stability based on key manipulation and hand position for using function
		Accuracy	Degree of accuracy of key manipulation by thumbs' movement
		Controllability	Degree of controllability of key manipulation for using functions
	Interference		Degree of interference between manipulated key and other keys
	Cognitive		Degree of cognitive key manipulation
	Performance		Degree of performing key manipulation task
Form-factor	Stability		Degree of grip-stability during key manipulation in each form-factor
	Conflict		Degree of conflict between key manipulation and form-factor's movement

3.3 Risk Level Definition

About PUI risk level which can evaluate PUI factors quantitatively was defined. Three values were defined. KMV in key-type is related to controllability and usability for performing task efficiently is defined. FMV of use-scene is related to controllability and usability of key manipulation for using functions. And about Form-factor, HV is related to controllability and usability of those is changed by transformation of mobile device's form. Table 5 shows the value which represents the degree of control efficiency for evaluating PUI risk level.

Table 5. Control efficiency value

Value	Requirement	Definition
KMV (Key Manipulation Value)	Key- type	Degree of control efficiency of each key-type during performing task
FMV (Function Manipulation Value)	Use-scene (Function)	Degree of control efficiency of each key-type during performing function
HV (Handling Value)	Form-factor	Degree of control efficiency of each key-type during performing function in each position

3.4 Risk Level Evaluation

PUI risk level is evaluated in 3 steps. Figure 2 shows evaluation procedure of PUI risk level. In first step, KMV is generated by evaluation of the degree of requirement's satisfaction in 5 investigated mobile devices' key-types. In second step, FMV is generated by evaluation of the degree of requirement's satisfaction in use-scene using KMV which is generated in first step. Similarly, HV is generated by evaluation of the degree of requirement's satisfaction in each form-factor using FMV. Lastly 1 minus HV value is risk level of mobile device.

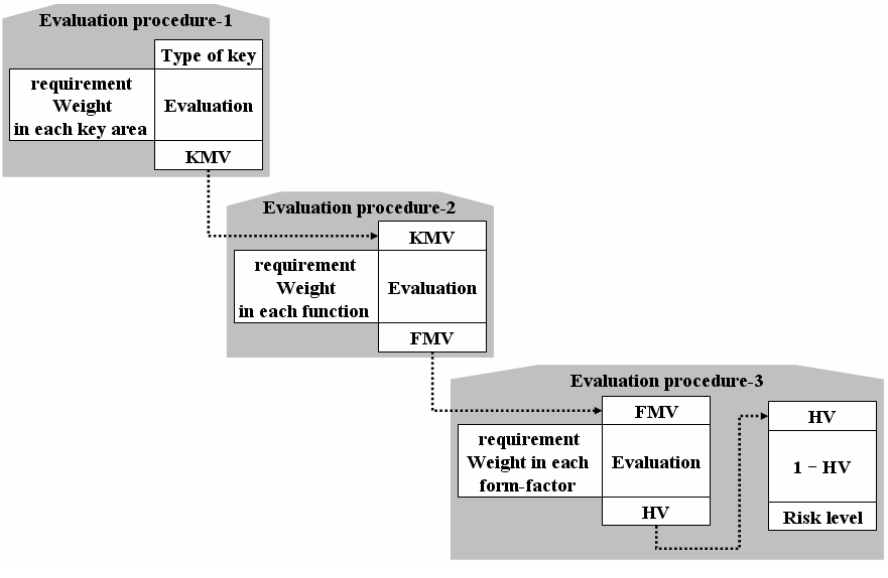


Fig. 2. Evaluation procedure of PUI risk level

Figure 3 shows the method that calculates KMV, FMV, and HV. Using this measurement method, for the last, PUI risk level was calculated.

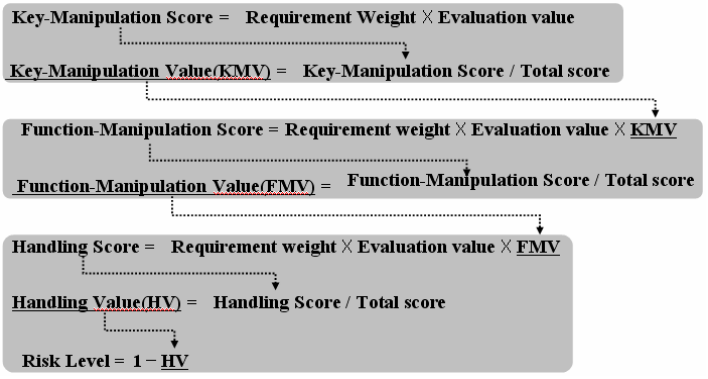


Fig. 3. Measurement method of KMV, FMV, HV

4 Conclusion and Discussion

The result of PUI risk level is produced by mobile device evaluation method of this paper. For example, figure 4 shows the results of risk level of 11 bar-type mobile devices.

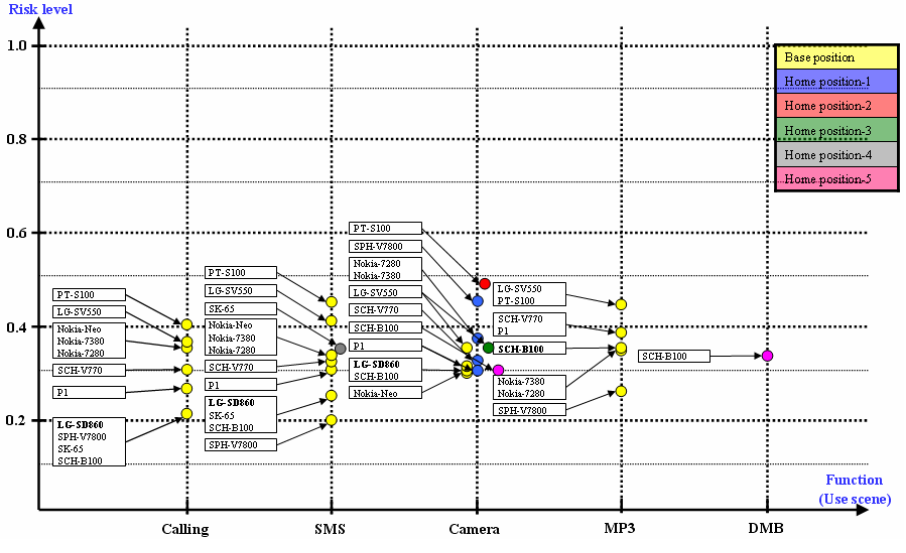


Fig. 4. PUI risk level of bar-type mobile device

In this paper, mobile device's requirements of PUI factors; key-type, use-scene and form-factor are extracted. And we developed risk level evaluation framework for mobile device's PUI design using quantitative value; KMV, FMV and HV. Using this framework, in early design step, designers can evaluate PUI risk level quantitatively.

In consequence, the framework provides quantitative result based on organized method for mobile device's PUI factors. Also as problems of PUI in early mobile device concept were predictive, designer can get an oppotunity to solve the problems easily. However, we only focused on PUI factors without GUI factors in this research so that the evaluation framework can't deal with whole part of mobile device. Therefore further work that includes menu structure and visual component(GUI factors) is needed.

References

1. Brewster, S.A.: Overcoming the Lack of Screen Space on Mobile Computers. Personal and Ubiquitous computing 6(3), 188–205 (2002)
2. Swindells, C., MacLean, K.E.: A Case-Study of Affect Measurement Tools for Physical User Interface Design. Graphic Interface, pp. 243–250 (2006)
3. Gilbert, B.C., Hahn, H.A., Gilmore, W.E., Schurman, D.L.: Thumb up:anthropometry of first finger. Human factors 20(6), 747–750 (1988)
4. How to Cope with Diffusion of Mobile Convergence. Samsung Economic Research Institute (2005) http://www.seri.org
5. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1994)
6. Lumsden, J., Brewster, S: A Paradigm Shift:Alternative Interaction Techniques for Use with Mobile & Wearable Devices (2003)
7. Murphy, J., Kjeldskov, J., Howard, S.: THE CONVERGED APPLIANCE – I LOVE IT...BUT I HATE IT. In: Proceedings of OZCHI 2005, Canberra, Australia (November 2005)
8. Kadefors, R., Areskoug, A., Dahlman, S., Kilbom, A., Sperling, I., Oester, J.: An approach ergonomics evaluation of the hand tools. Applied Ergonomics 24(3), 203–211 (1993)
9. Hornbaek, K.: Current practice in measuring usability-Challenges to usability studies and research. Human Computer Studies 64, 79–102 (2005)
10. Hirotaka, N.: Reassessing Current Cell Phone Designs-Using Thumb Input Effectively. CHI 2003, New Horizons (2003)
11. Ha, R.W., Ho, P.-H., Shen, X.S.: SIMKEYS-An Efficient Keypad Configuration for Mobile Communications. IEEE Communications Magazine, University of Waterloo, (November 2004)
12. Shneiderman, B.: Designing the User Interface Strategies for Effective Human Computer Interractive Working. Addison-wesley, Reading (1987)
13. Silfverberg, M., Mackenzie, I.S., Korhonen, P.: Predicting text entry speed mobile phone. CHI2000 Letters 2(1), 9–16 (2000)

Tracing Cognitive Processes for Usability Evaluation: A Cross Cultural Mind Tape Study

Jyoti Kumar[1], Janni Nielsen[2], and Pradeep Yammiyavar[1]

[1] Department of Design, IIT Guwahati, India
[2] Department of Informatics, CBS, Copenhagen, Denmark
{jyoti.k,pradeep}@iitg.ernet.in, jn.caict@cbs.dk

Abstract. Cultural differences in cognitive processes and cognitive tools have been extensively documented. Design and use of culturally sensitive interfaces have been in demand in HCI for sometime. In this study the method of stimulated retrospective verbalization which is called here as Mind Tape study, has been used to capture cognitive differences of Danish and Indian users while interacting with chosen websites on a given task. The recording of the interaction captures screen activities and video of user. The replay of the recording is used as stimulus during a voice over interview. Using Mind tape, not only the sequence of activities during task fulfillment is observed, but also an insight into the user's cognitive processes, motives and intentions, regarding the choices made and activities done are recorded. The paper reports the cultural sensitivity and suitability of the mind tape method for cross cultural usability evaluations in light of the study conducted.

Keywords: Stimulated Retrospective Verbalisation, Usability testing, Cross Cultural.

1 Introduction

Verbalisation as a window to the cognitive processes of the user has been a well talked of method in the usability evaluation practice [1,2,3]. Concurrent and retrospective verbalisations have been compared and contrasted for their reliability and validity often [4,5,6]. Whereas the Concurrent Verbalisation (CV) suffers from sharing the cognitive resources with task fulfillment [3], the retrospective verbalisation has been accused of memory loss due to time lag or subsequent influences on STM. The validity of Stimulated Retrospective Verbalisation (SRV) or Mind Tape (for under the influence of the stimulus mind acts as a tape and unwinds the memory thread by thread) have been established by a few studies [7,8] and the quality of Mind Tape data also have been reported as compared to CV like Think Aloud (TA) [7]. On the other hand cultural differences in social setups have been reported [9] and cognitive basis of the cultural differences have also been argued [10]. Now the issue of culturally sensitive methods of usability evaluation is being raised in this paper. When there exist cognitive differences in cultures, do we also need to examine the suitability and sensitivity of usability valuation methods in cultural and

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 336–345, 2007.
© Springer-Verlag Berlin Heidelberg 2007

cross cultural contexts? In this study mind tape has been used as a method to examine cognitive processes cross culturally, namely with Danish and Indian participants, over a task of exploration and finding a place of interest in each country of interest to visit using national tourism websites of three countries, India, Denmark and China. The results suggest that Mind tape gives rich data for analyzing the cognitive processes and tools employed by the users in task fulfillment and the method is culturally suitable to both the cultures in terms of satisfaction reported and data gathered.

2 Method

2.1 Website Explorations

Websites of three countries were selected. The sites are the official tourist websites of the three countries. They all address the same target groups; potential tourists, and English version of all of them were available. The user studies were conducted with 7 pairs (a user and an interviewer) each from Copenhagen Business School, Denmark and from the Indian Institute of Technology Guwahati. A scenario framed the task which was I) to explore the three sites and II) to find a place of interest in the tourism websites to take back to a group of friends planning a holiday trip. All seven user/interviewer pairs came from interdisciplinary studies where computer science made up one part combined with another discipline. In Denmark students came from Copenhagen Business School, Department of Informatics. They were master students (beginning of 4th year) and had enrolled in a course in HCI. They had some idea of interface design and evaluation aspects. The students from India were bachelor students in their final 3rd year at Department of Design, IITG and they had a similar educational background in design and evaluation of interfaces. A few had visited the website of their home country, but none had visited or were familiar with all three websites. None had explored the sites extensively as they were requested to do during the task. The mean age of Indian participants was 21.57 with standard deviation of 0.73, the mean age of Danish participants was 26.14 with standard deviation of 2.29.

Table 1. Profile of Danish Participants

d	Male	Female	Age
d1	1		28
d2	1		26
d3		1	28
d4	1		26
d5	1		24
d6	1		29
d7	1		22
	Total 6	Total 1	Mean 26.14
			SD 2.29

Table 2. Profile of Indian Participants

	Male	Female	Age
I1	1		23
I2	1		22
I3	1		21
I4	1		21
I5	1		22
I6	1		21
I7	1		21
	Total 7	Total 0	Mean 21.57
			SD 0.73

The users were asked to correlate between cursor and the user's eye as they browsed through the websites for the task fulfillment. To enhance this correlation the users went through a training session to learn to coordinate cursor and eye. During the first task, the exploration, the users were encouraged to get a feel for the country so as to be able to communicate to her/his friends. The second task was to find one place of interest in a website where the user would like to go with her/his friends for a weekend. To get around the problem with verbal overshadowing of TA, and to allow the visual interaction to unfold undisturbed, no requests for concurrent verbalization were made. The user worked at her/his pace and in peace during the whole session.

The data was collected by recording the entire interaction on the screen including a video image of the user. Immediately following each task, the interviewer replayed the recording and conducted a qualitative interview. The software used made it possible to record voice over the original recording. The interviewer paused and played the original screen recording asking the user questions like for e.g. "what are you looking for?" when the user's mouse is seen wandering around on the screen for sometime without clicking, or "Why do you click there?" when the user clicks at some link. The answers from the users were developed upon to further probe into the user's intentions and expectations. A questionnaire was applied at the end of all the three website explorations to get additional information about the overall view of the websites and the experiences with mouse eye coordination.

3 Results

3.1 Mind Tape Study

The Interactions with three websites were screen recorded and voice over interviews were conducted on them. Finally the voice over video was analysed for the users' responses to the interviewers' questions regarding what they were doing at specific instances during the website explorations. Some of the noteworthy observations are listed subject wise in Table 3 as an example of the kind of data that was obtained from the Mind Tape study.

Table 3. Observations from Mind Tape video

Subject	Observation	User's response	Inference
d1	Indian Site: Mouse wanders in the beginning, checks the menus.	Looking for 'Tajmahal' for I have heard of only that from India.	Posit: Danish People/ In general people search by what they already know on an unfamiliar website
	Clicks 'Heritage' link	Expected that it will give me some pictures of Tajmahal.	Pictures are what information can be quickly and richly availed.
	Picture of Tajmahal comes on screen. Expression of dissatisfaction on the face of subject.	Got only one picture with little text so I started looking for some other link where I can get more info.	Need of many pictures.

Table 3. (*Continued*)

Subject	Observation	User's response	Inference
	Text of info comes	I am looking for pictures, I am not going to read 10 pages of text.	Lot of Texts is not preferred on a tourism website.
	Danish Site: Beginning… mouse wanders around	I am looking for something interesting	Posit: When the person is well aware of the place then one looks for something interesting (does it mean not known earlier)!
	Clicks link named '*inspiration*'	I am looking for something interesting so I guess here is something…	The word 'inspiration' promises for new and exciting on a tourist website.
	Further sub menus come upon clicking *inspiration* - sub menu - *culture*	I am not looking for so specific information when I click culture, I want a general picture.	There is a threshold of detailed information that one seeks while looking for a tourist place, at least initially.
	Chinese site: Beginning… Mouse wanders…	I thought Hong Kong is part of China, I am not able to get it.	Again search by know place on a less known site.
d2	Indian Site: Mouse static in the beginning.	Looking for some pictures to see what all places to visit in India, I do not know much about India.	Pictures as means of getting an image of the place.
	Selects *Beaches of India - Goa*	Because it has pics of beaches so I can go there	Probably familiar locations interest more
d3	Selects *places to visit*	I do not know anything about India so may be this is a good place to begin with.	Cognitive tools that aid in beginning to search is not names of places for they are unknown but the categories that represent them. This could possibly be a universal phenomenon.
	a list of places is shown	I do not know any of the places so this list doesn't give me desired information.	Further categories of places and then the list might have helped probably.
I1	After a lot of trials on menu items	The purpose of this website is not clear… whether it is about introducing me to the culture…. Or it is also to help me get there…	Could it be much talked about- holistic thinking in east Asians…. Trying to get the bigger picture?

Table 3. (*Continued*)

Subject	Observation	User's response	Inference
I2	Looks at an image	It looks like from my very own place	Does this cultural identity phenomena relevant more to this individual or to the community?
I4	Gets a submenu filled with known items except one	These I know…OK… but what is this?.. let me click	Posit: In known territories, people explore the less known to them item.
I5	State wise organization of info	Why is it done state wise? I am interested not in states but the kind of holiday I want to have.	Information architecture to suit the motivation of the user was observed in user's of both the cultures.
I6	Highlights the text while reading	I always do it while reading, it helps me identify the text from rest	Cognitive tool used by most of the Indian participants while reading to focus on the text being read. Is it a cultural phenomena?

3.2 Rankings of the Websites by the Subjects Under Different Criteria

The subjects were asked to rank the websites after the task fulfillment was over. The criteria given were 'the website they liked', 'The website that was most easy to use' and 'the website which had most pleasing interface'. Table 5 lists the frequency of ranks allotted to the websites by the Indian, denoted by 'Ind' and Danish user's denoted by 'Dan' under each criteria given for every website.

Table 4. No. of participants from India (Ind) and Denmark (Dan) who ranked the sites under the criteria of liked, easy to use and Good Interface

	Indian Website						Danish Website						Chinese Website					
	Liked		Easy to use		Interface		Liked		Easy to use		Interface		Liked		Easy to use		Interface	
Rank	Ind	Dan	Ind	Dan	Ind	Dan	Ind	Dan	Ind	Dan	Ind	Dan	Ind	Dan	Ind	Dan	Ind	Dan
1st	**5**	**3**	**4**	**2**	**2**	**1**	**2**	**3**	**1**	**4**	**5**	**5**	**0**	**1**	**1**	**1**	**0**	**1**
2nd	1	2	2	2	4	4	4	3	2	3	2	1	2	2	2	2	1	3
3rd	1	2	1	3	1	2	1	1	4	0	0	1	5	4	4	4	6	3

3.3 Grading of the Websites by the Subjects

After the task fulfillment, the subjects were also asked to rate the websites on a 7 point scale for how 'attractive to look', 'exciting to visit' and 'friendly to use' each website was. The results have been tabulated in Table 5 with mean (with standard deviation), maximum rating and minimum rating that each website got from Indian (Ind) an Danish(Dan) participants.

Table 5. Mean of ratings of three sites on a 7 point scale under criteria of Attractive, Exciting and Friendly of Indian (Ind) and Danish (Dan) participants

	Indian Website						Danish Website						Chinese Website					
	Attractive		Exciting		Friendly		Attractive		Exciting		Friendly		Attractive		Exciting		Friendly	
Rank	Ind	Da	Ind	Da	Ind	Da	Ind	Da	Ind	Da	Ind	Da	Ind	Da	Ind	Da	Ind	Da
Mean	**5.4**	**5.0**	**5.4**	**3.6**	**4.5**	**4.5**	**4.7**	**4.3**	**4.0**	**4.0**	**4.7**	**5.0**	**3.5**	**3.8**	**3.4**	**3.6**	**3.2**	**3.1**
Std. Deviation	.7	1.4	1.2	1.2	1.5	.8	.9	1.2	1.0	1.0	1.7	.8	1.1	1.7	1.1	1.8	1.1	1.3
Minimum	5	3	4	2	2	3	3	2	2	2	1	4	2	2	2	1	2	2
Maximum	7	7	7	5	6	5	6	5	5	5	6	6	5	6	5	6	5	5

4 Discussion and Conclusion

4.1 Hand Eye Coordination

3 Danish participants reported 'no problem' using the hand eye coordination and that is was 'natural', 2 Danish participants reported that is was difficult when 'scanning the pages' and that 'the eye moves faster than the hand'. 4 of Indian participants reported the difficulty in 'scanning the page' and 2 said 'it was natural while reading as one always does that.

Inference: Hand eye coordination as a means to get the data about visual focus of attention on the screen may be natural to some and they otherwise also may have a tendency to take the mouse where their eyes went in normal interactions. Whereas, to some, it was intrusive in their normal task fulfilment activity. Anyhow, text reading was observed to involve the cursor movement along with the text being read naturally, many Indian participants had shown the behaviour of highlighting the text being read for better attention. It may be posited here that the mouse track data can be a good approximation to the eye gaze data in case of the reading activities like text reading, menu items reading etc. but it may not be reliable in case of image viewing or scanning webpages.

4.2 Quality of Mind Tape Verbalisation

Subjects were able to recall satisfactorily what they were thinking/doing at the time of the activity being replayed on the screen. The voice over interviews yielded considerable data on the why's and how's of the activity. The Indian participants specially, gave an extended set of logical explanations of what made them do the activity, some even presenting their views about the site in general while the activity was being interviewed. Participants from both the cultures were comfortable in the mind tape study and the information related to cognitive processes and tools applied for task fulfilment were satisfactorily reported.

Inference: The rich set of verbal data corresponding to each activity which was possible in mind tape study could have possibly interfered with their normal task fulfilment in the concurrent verbalisation. The satisfaction that was seen in the users sharing their *why's* and *how's* of activity due to a human angle to the verbalisation namely, the interviewer, could have possibly not been there in the monotonous concurrent verbalisation. This helped in getting deeper insights into the cognitive processes employed by the users. In general it can be said that mind tape is a culturally sensitive tool for usability evaluation tasks.

4.3 Cross-Cultural Similarity in Cognitive Processes Employed

Search by familiarity: All participants, when searching in little known countries to them (like Denmark and China in case of Indian participants and vice versa in case of Danish participants) ordered their search from more known to less known places. Like in case of Danish participants exploring Indian website, 3 out of 7 participants, started their search from looking for 'Tajmahal', which was the only place in their prior knowledge (as reported), but upon finding no images corresponding to Tajmahal, moved to what interests them, like some of them searched for beaches in India. Whereas subjects looking for places when confronted with known set of places, looked for the one that was little known to them. But in finally deciding about the places, people based their decision on the combination of prior knowledge and supplements of info from the website.

Inference: The search by familiar could be phenomena common to the two cultures, or it could very well be universal phenomena, in case of travel websites. On the other hand the inquisitiveness for the odd one in the list of known places could either be an attempt for the mere information sake.

Role of images in decision making: Almost all participants from Denmark as well as India complained for the lack of pictures in Indian website. They articulated the role of images in getting a feel of the place to visit. They also closed very quickly those pages of the site that had no images.

Inference: This may speak of the similarity in the cognitive processes and tools of the users from both the cultures or it may be a universal phenomena. This needs to be further investigated.

4.4 Cross-Cultural Similarity in Cognitive Processes Employed

Query of cost as an aid in decision making in Indian participants: 5 out of 7 Indian participants searched for the prices of the facilities and used the information as a

primary aid in deciding about the places to visit. This behaviour was observed only in 2 of the Danish participants though.

Inference: As the sample under study is very small to generalise the inference statistically, still it becomes a significant suggestion towards further exploration into how do people from the two cultures employ cognitive tools in decision making.

Online reading habits: Indian participants (5 of 7) were found to select the text with mouse as they read, in the mind tape study they reported it as their normal habit while reading. None of the Danish participants had this habit.

Inference: Could it be possibly due to differences in cognitive tools people employ while seeking information online, specially through reading? The holistic verses analytical cognitive processes (in East Asians and westerners respectively) reported by Nisbet et al [1] are in action here? Further specifically designed experiment to study this phenomenon in web based information seeking behaviour could be conducted to verify/substantiate it.

4.5 Rankings of the Websites *(Table 4)*

As depicted in the table, 5 of the 7 Indian participants liked the Indian website most and said that they found it 'Organised', 'had images with relevant info', 'Concise with important guides', 'had Nice colours' and 'had relevant chunking of information' while 3 Danish participants liked the Indian website (giving reasons 'not confusing', 'information was grouped well') the most and 3 liked the Danish website the most (giving reasons, 'easier to navigate', 'had light colours' and 'was structured'). The one Danish participant who liked the Chinese website gave the reason of it having a lot of pictures. 4 of Indian subjects found the Indian website most easy to use (reasons elicited were 'front page had sufficient information, 'grouping of info was good', 'could locate places more easily' etc.) while 4 of Danish participants found the Danish website most easy to use (because of 'lots of useable links on the front page', 'Clear separation of information', 'menu made it easy' etc.). Importantly, two of the Indian participants who had liked the Indian website most and the Chinese website the least found the Chinese website most easy to use for it 'had a linear structure' and 'had nothing to search'. Both the Indian and the Danish participants (5 in both) found the Danish website having most pleasing to Interface an the reasons given were 'had best colour codes', 'had good selection of fonts, colours, photographs', 'had a neat and clean look' (Indian participants), 'had simplicity', 'was clean' (Danish participants).

Inference: a) Both the Danish and the Indian participants found the Danish website clean and simple which can help us hypothesise that the cognition of neatness and cleanliness depend on similar visual cues in both Danish and Indian cultures. b) A lot of pictures (and only pictures in eyes of d4, I4 and I5) on the Chinese website helped in deciding about the place in that country but it couldn't win for the most liked website for both groups of participants, from this we could hypothesise that though pictures become the most important element for deciding about places in tourism websites for the users of Danish and Indian origins (as reported by both groups of participants), but it doesn't win the site most liked award for user's of both countries prefer 'organisation' and 'neatness'. c) The ease in use reported by both groups for

their own culture website may owe to either the familiarity of the information available on the website or it may be due to cognitive difference the user groups have in reality. Further study need to be conducted for validation.

4.6 Ratings of the Websites *(Table 5)*

Indian website was reported as most attractive, getting an average rating of 5.4 by Indian participants and 5.0 by Danish participants on a 7 point scale whereas the Danish website was reported to be more friendly to use by both the groups (4.7 and 5.0 respectively). The divide in opinion of the two groups in terms of which was the most exciting to visit is clear when Indian group has favoured it's native site (5.4) in comparison to Danish site (4.0) whereas the Danish subjects have rated both the sites almost similarly (3.6 and 4.0) respectively.

Inference: The Indian site appearing attractive could possibly be attributed to it's bright orange colour in the layout, and plethora of selected images on the header which though one Danish participants disliked and many Indian participants ranked the Danish site's interface more clean and with soothing light colours, still during rating the Indian site has got more points under attractive attribute. Low ratings of Chinese website may be due to the unobvious position of links (which was on the images). Those subjects (like d4) who could figure out the links have rated it relatively higher because they could find a source to a lot of images which has been reported as one of the most sought after information sources.

To conclude, mind tape study does reveal insights into the cognitive processes of the users by developing upon and probing into the user's responses to the questions related to the activity they had just finished during task fulfilment. Furthermore, the human angle in the form of interviewer makes it easier and more meaningful to have a dialogue about the intentions and motives of the user in employing the cognitive tools, in form of choices, aids in decision making, preferences for colours, forms, images etc. while they perform tasks. The cultural suitability of this method to the two cultures under study has also thus been established. The study has revealed the common cognitive tools in two cultures like 'search by familiarity' and 'role of images in decision making', 'similarity in concept of clean and neat site' and the differences in form of the 'online reading habits' and 'search for holistic impressions'. The mental models of 'attractive site' and 'friendly site' have also been found to match for the two cultures. This study also advocates for the cultural sensitivity of this method as dialogue oriented cultures, it is posited here that Indian culture is one (on the basis of *relational, dialectical* and *person attribution* in Peng K [11]), as well as task oriented cultures, it is again posited here that western cultures are task oriented (on the basis of *non-contradictional* and *event attribution* in Peng K [11]), will find it suitable to have an interviewer to speak out their motives than just one way, monotonous verbalisations as in case of concurrent verbalisations like Think Aloud etc. These findings could help formulate further studies using mind tape method to explore the cross cultural cognitive process and tools differences and similarities in more detail.

5 Future Work

Further studies in finding role of the interviewer in the elicited data, impact of cross cultural pairs in voice over interviews for usability testing, probes useful in specific cultures, whether there are culture specific probes, could be conducted to expand and explore the possibilities of application and validity of mind tape method. The collected data itself is being further analysed for cultural cues for the method.

Acknowledgements. This study was co-funded by the Danish Council for Independent Research (DCIR) through its support of the Cultural Usability project.

References

1. Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and Systems of Thought: Holistic Versus Anallytic Cognitoin. Psychological Review 108(2), 291–310 (2001)
2. Nielsen, J.: Usability Engineering. Cambridge, MA, AP Professional (1993)
3. Ericsson, K., Simon, H.: Protocol Analysis – Verbal Reports as Data. MIT, Cambridge (1993)
4. Boren, M.T., Ramey, J.: Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication 43(3), 261–278 (2000)
5. Nielsen, J., Clemmensen, T., Yssing, C.: Getting access to what goes on in people's heads? - Reflections on the think-aloud technique, Nordi CHI (2002)
6. Nielsen, J.: Reflections on concurrent and retrospective user testing. In: Proceeding of India Human Computer Interaction Confr., Bangalore (December 2004)
7. Van den Haak, M., De Jong, M., Jan Schellens, P.: Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour and Information Technology 22(5), 339–351 (2003)
8. Zhiwei, G., Shirley, L., Cuddihy, E., Judith, R.: The Validity of the Stimulated Retrospective Think- Aloud Method as measured by Eye Tracking, CHI 2006 proceedings (2006)
9. Geert, H.: Cultures consequences, 2nd edn., Sage publications
10. Nisbett Richard, E.: Geography of thought, Free press
11. Peng, K., Ames, D.R., Knowles, E.D.: Handbook of Cross-Cultural Psychology. Oxford University Press, Oxford (2000)

Lessons from Applying Usability Engineering to Fast-Paced Product Development Organizations

Dong-Seok Lee[1] and Young-Hwan Pan[2]

[1] Software and Solution Center, CTO, LG Electronics,
679 Yeoksam-dong, Gangnam-gu, Seoul 135-985, South Korea
dongseok.lee@lge.com
[2] Graduate School of Techno Design, Kookmin University,
861-1 Jeongneung-dong, Seongbuk-gu, Seoul 136-701, South Korea
peterpan@kookmin.ac.kr

Abstract. This study discusses why usability engineering, which seems easy to contribute to more usable products, finds little support in fast-paced product development organizations. It discusses the ways in which the environment of a product development organization is quite different from that of a web or software company. Among the differences are faster-paced development, more rigorous process stages, lower number of iterations allowed, and higher cost for usability amendment. Thus many usability professionals cannot escape from the traps of simply fixing glitches instead of solving major problems, and working on product issues only in reaction to usability problems generated by other stakeholders. This study provides some innovative suggestions for usability professionals as effective alternatives to remaining stuck in the typical evaluation and refinement strategy of usability engineering.

Keywords: Usability engineering, product development organizations, limitation of usability engineering, ROI of usability engineering.

1 Introduction

Usability is now widely accepted as an important part of product development in many companies. Usability professionals participate in ongoing development projects, evaluate the usability of developing products, and suggest design alternatives, by which they contribute to the development of more usable and satisfying products. Many practices have reported success stories of usability engineering, which describe how usability engineering could contribute to better usability, increased productivity, and/or reduced cost of development and service [1, 2, 3].

Can these stories be applied to a product development organization? It is observed that most of the success stories have come from the domain of website and software development. A very small number of usability success stories come from product development organizations such as Nokia, Sony, LG, Samsung, or Philips. The ultimate goal of usability professionals in an organization is to make their companies into "user-driven corporations," as Nielson termed it [4]. When we locate some of the

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 346–354, 2007.
© Springer-Verlag Berlin Heidelberg 2007

well-known companies in Nielson's corporate usability ranking, we can easily see the tendency for product development organizations to have ranked lower than software and web companies. The maturity level of most product development organizations would be lower than stage 5, but the level of some software and web companies such as Google, Microsoft, EBay, and Amazon goes beyond stage 5.

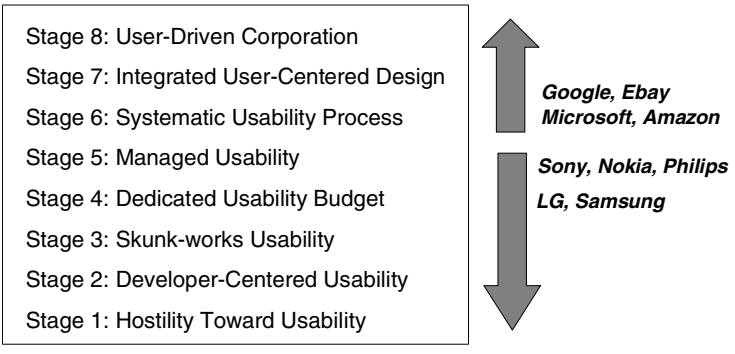


Fig. 1. Corporate usability maturity level of some companies (Subjective judgment by the authors based on known facts about whether a usability team exists in the company, and its role in the development process)

This study will discuss why it is difficult for usability engineering to succeed in product development organizations – especially fast-paced development organizations. The authors will present the lessons they have learned through experience at product development organizations, and will suggest the strategy of usability engineering.

2 Chasing Usability

> *"Spending $60,000 on usability engineering throughout development resulted in savings of $6,000,000 in the first year alone"* [5]

As pointed out in many usability-related studies [4, 6, 7], forming a usability team is the first organizational milestone for usability professionals. The next step would be usability institutionalization by which usability evaluation becomes a set of mandatory steps in the product development process. Senior managers tend to expect that forming a usability team and promoting usability institutionalization will contribute to increased usability and better user satisfaction. Are the managers' expectations realistic? Here's what happened to the authors in a fast-paced product development company.

With strong support from the executives, the company organized a usability team in order to increase user satisfaction and restrain increasing service costs. The usability team scouted usability professionals, built a testing laboratory, and revised the development process by making usability evaluation a mandatory part of it. Figure 2 shows how the usability team collaborated in the product development process. The

team participated in all of the major development projects beginning with the requirement phase where product planning occurs, and evaluated usability through heuristic evaluation, cognitive walkthrough, and laboratory testing, both at the conceptual design stage and at the development stage. The team was permitted to force designers and developers to solve the usability problems found. The team also educated developers about usability and its importance to meeting business goals and objectives.

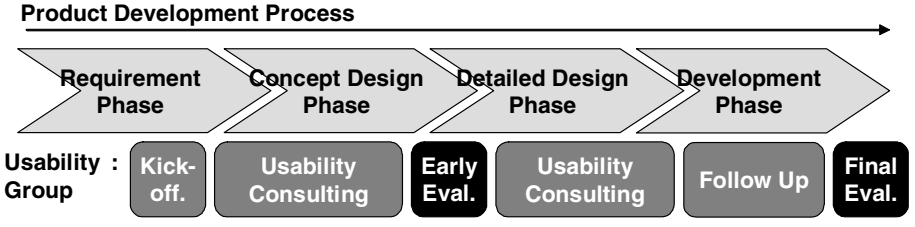


Fig. 2. Usability evaluation stages in the product development process. Two mandatory usability evaluations: the early evaluation after the conceptual design stage and the final evaluation after the product implementation stage.

In their experience, the authors found that the environment at product development organizations was quite different from that of software and web companies. First, the stages of the product development process were faster and more rigorous because of the higher pressure from time-to-market. This means that only a small number of iterations of design and evaluation were allowed, depending on circumstances. Thus the usability team had to expedite usability evaluation as quickly as possible in order to avoid being a delaying factor. Second, the development process was almost irreversible; the cost of solving usability problems skyrockets at later development stages. The team was forced to find and fix usability problems as early as possible. Third, more usability-related groups exist in a product development organization. Among them are design, software, hardware, product planning, marketing, sales, and quality assurance. Thus resolving usability problems requires more complicated and sophisticated collaboration and intervention.

Despite the difficulties presented above, it appeared that the usability team succeeded in meeting the managers' expectations. They solved many usability problems, helped to re-define the role of user interface groups, increased the awareness of usability, and piqued the senior managers' interest.

After two years, however, it was reported that the return rate and service costs of most of the company's products had continued to increase similarly to those of other competitors, and some products' return rate reached almost 40% [8, 9, 10]. The team members were puzzled at the unexpected results, and the managers started to doubt the value of the team. The team tried to prove the value of usability engineering but they could not show persuasive data to the managers; reduced service costs were not significant enough and developers viewed usability evaluation steps as a delaying factor.

3 Challenges to Usability Engineering

> *"You say that we have solved a lot of usability problems and the usability level of our products has been increased, so how is it that there's no sign from the market that we have a better product than before? When do we start to see a return on our investment in usability?" The manager of the usability team (translated from Korean).*

Eventually, the company downsized the usability team, relocated some of the members to user interface development groups, and moved the usability engineering role to the Quality Assurance group. Although this is the experience of the authors, the same story takes place in some other companies.

What was wrong with the team? What kept them from being successful in proving the value of usability engineering? Existing studies [6, 7, 11, 12] may show that the team had poor usability evaluation skills, were managed by the wrong person, were in the wrong place in the organizational structure, chose inappropriate testing methods, wrote poor-quality reports, caused delays in the market launch, and/or failed to build good relationships with others. Any of these factors may have influenced the managers' decision to discontinue the operation of the usability engineering group.

But can we apply the above issues to the usability teams of Microsoft, Yahoo!, Sony, Nokia, and Philips? Those companies have very competitive usability teams. For example, Microsoft is known to have active usability teams with superior manpower, facilities, and authority to fix problems. Because other usability books have covered those issues well, this study will not discuss them. Instead, we will discuss the problem from a different perspective – a larger and broader perspective on design management.

A post-mortem analysis found four limitations of usability engineering in fast-paced product development organizations.

Fixing Glitches: First, the usability team had focused on solving glitches rather than serious complexity problems. Solving complexity problems requires longer design exploration and smart, solid ideas; thus, a fast-paced developmental environment prevents the team from solving complexity problems [13]. For example, a series of usability evaluations of a cell phone user interface found that users had difficulty finding appropriate menu commands. This was not an easy problem to solve, since most cell phones today have hundreds of menu commands. Even restructuring menu organization doesn't guarantee better usability. Consequently, the usability team was forced to postpone finding resolutions to the complexity problems, electing instead to focus on smaller problems that they could solve quickly (this is known as "glitch repair").

Reactive: Second, the team was working reactively; their development work consisted only of responses to the new usability problems generated by the added new features. Indeed, electronic products evolve quickly, continually sporting new features, and with each set of new features comes a brand new set of usability problems. The usability professionals, in their reactive position, were just trying to keep up. The role of usability professionals in an organization is to eliminate usability problems [13]. But

complexity increases exponentially – not linearly – in proportion to the number of features. As Norman [14] noted, complexity seems to increase as the number of features squared. This means complexity increases at a faster rate than the rate at which usability engineering can fix it. Therefore once complexity has started to dominate, investing in usability doesn't guarantee better products.

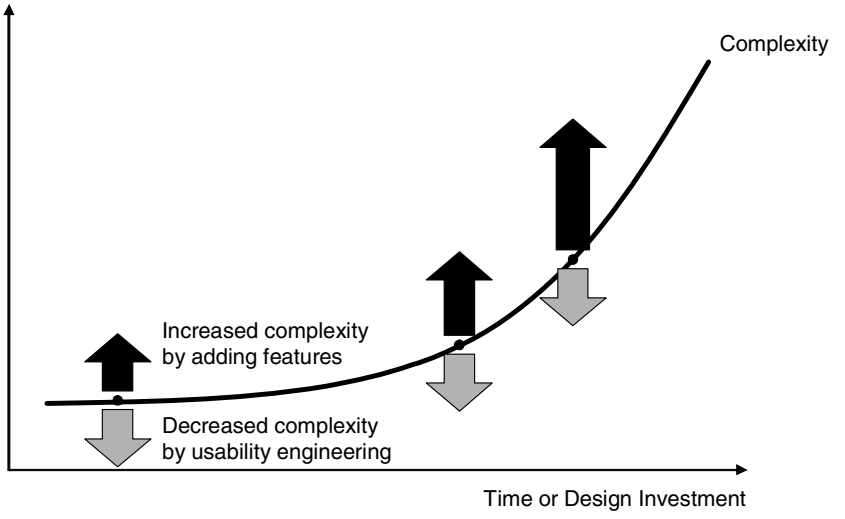


Fig. 3. Complexity curve in feature-overloaded products: Complexity increases exponentially as features are added. Over time, complexity increases at a faster rate than the rate at which usability engineering can fix it.

Lower financial benefits: Third, the financial benefit of usability engineering was not as high as described in the success stories for web and software companies [1, 2, 3]. Unlike with web companies, the reduction in service costs at electronics companies was not found to be high enough to justify usability engineering. For example, if a usability team reduced ten thousand service calls (10 percent of total service calls – a huge achievement in product development organizations), then the amount of savings in service costs was fifteen thousand dollars (if we assume $1.5 service cost per call). This amount of money wasn't sufficient to cost-justify usability investment. In addition, the development group insisted that usability evaluation caused a longer development time. Even worse was the overall trend of increasing service costs and a return rate that had not diminished.

Lower impact on point-of-purchase: Fourth, usability is a less critical buying factor for electronics products. People do like products with more features and more buttons, even if they look complicated [15]. While people use and pay for websites, people pay for and use electronics. And once they have paid, they use the product for a longer period of time. Thus better usability has been politically defeated by faster development, better product design, and lower price.

4 Lessons and Suggestions

The authors interviewed several usability professionals from other product development organizations, and collected their stories. Almost all of them agreed on the limitations of usability engineering in fast-paced product development organizations, and said they were looking for alternatives to overcome the limitations in order to survive in their companies. In the meantime, the authors heard that one usability team disbanded because of its inability to solve major usability problems. It seems that evaluation and refinement, the normal strategy of usability engineering, doesn't work very well in fast-paced product development organizations.

The authors' experience and discussions with practitioners gave the following lessons regarding applying usability engineering in fast-paced product development organizations.

Invest more in solving major usability problems: First, usability professionals should focus more on solving complexity problems. Dependency on the typical strategy of evaluation and refinement, combined with fast-paced development schedules, forces practitioners to focus only on solving glitches. Usability professionals should avoid this trap by leveraging usability evaluation activities and design exploration activities. In many cases, major usability problems such as too much navigational complexity on cell phones and digital television menus, are common problems for competitors. This means that these problems are difficult to solve, but solving them could easily result in higher market competition. We call this *user interface innovation* – in the same fashion as Apple iPod's click wheel and Sony's cross media bar. Leading user interface innovation by solving usability problems is the key to flourishing usability.

Be proactive to future usability problems: Second, usability professionals should anticipate the usability problems of future products. Again, once complexity has begun to dominate and diminish products' value it is very hard to solve usability problems within a given amount of time. Despite a fast-changing environment, companies have development plans spanning a couple of years that describe what features are to be added in the future. Based on these plans, usability professionals should anticipate possible usability problems, and be prepared to solve the problems. For example, usability professionals of DVD players should ask what would happen if, with current information architecture, a DVD player had the functionality of downloading movies from a website and recording 25GByte blue-ray.

Measure the loss of investment: Third, return on investment (ROI) of usability in product development organizations should be calculated based on the loss of investment, such as development costs or loss of customer value. Reducing service costs and development costs seems unpromising when products continue to become more complicated as more features are added. It is also unhopeful to measure increased customer value by usability [8], since too many aspects of the products, such as price, brand, promotion, and market competition, affect the customer value. An alternative is to measure loss of investment. When products are in the planning stage, we expect that the value of the products is the sum of the capability of each feature. But after user interface design and development, the real value of the product decreases because of poor usability, which can be calculated by multiplying a product's usability by its capability (Figure 4). Thus, if we can figure out the cost for developing the

capability, the loss of investment by bad usability can be calculated by multiplying (1 – usability) and development cost. For example, if a company invested ten million dollars for development and the product's usability index was 0.8, then the loss of investment is two million dollars. If the usability team increases the index to 0.9 then the value of usability improvement is one million dollars. Since product development organizations invest large amounts of money to develop products, this approach to measure ROI of usability makes sense even during the process of development.

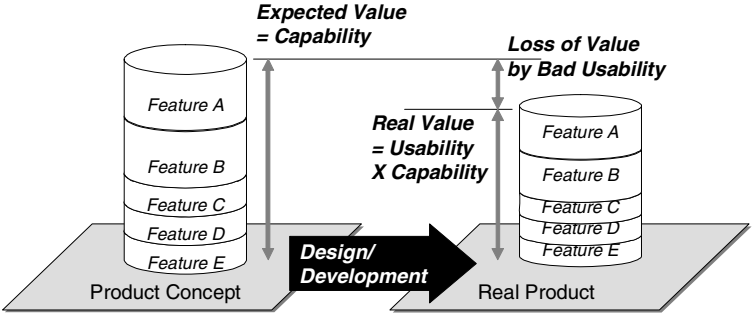


Fig. 4. Expected customer value vs. real customer value of a product: Product's value became smaller because of poor usability after design and development

Re-positioning usability: Finally, the role of usability teams should be re-defined so as to survive in organizations. The strategy of evaluation and refinement doesn't guarantee more usable products in a fast-paced development environment where refining a user interface is highly restrained. Lee at al. [13] refer to usability engineering as "following standard prescriptive approaches to technology development by solving short-term small repair problems at a local scale, and usually having to swim upstream relative to pressure on design, thus being trapped in a narrowing solution space." Beyond swimming upstream, a usability team needs to gather user stories, observe users' exploits and workaround, anticipate future usability problems, and provide cues for user interface innovation. The tendency of changing the name "usability team" to "user experience (UX) team" in many companies supports the authors' viewpoints.

5 Conclusion

This study discusses why usability engineering that seems so easy to apply to product development organizations is so difficult to achieve in practice. Based on users' experiences and interviews with usability professionals, the authors suggest that the environment of product development organizations with fast-paced product development schedules, new features, higher pressure of time-to-market, more rigorous product development processes, and a lower number of iterations allowed are quite different from that of web or software companies, where usability engineering has been applied and studied with more emphasis. Having practiced usability engineering for two years in a product development organization, the authors found that they and

their colleagues were trapped in fixing glitches and being reactive to usability problems, and failed to cost-justify usability by calculating reduced cost for service and development. It seems that existing usability engineering studies do not take into consideration product development organizations such as consumer electronics and healthcare products. This paper provides some organization, operation, and strategic suggestions for usability professionals in product development organizations. The authors hope that this study helps other usability professionals to survive and flourish in an organization, and trigger studies for usability engineering in product development organizations.

Acknowledgement. Prepared, in part, through IT Scholarship of IITA, the Ministry of Information and Communication of the Republic of Korea. The authors appreciated all the usability evaluation team members (alphabet order) – Ms. Hee-Jung Choo, Ms. Yerjean Han, Ms. Eun-Young Im, Dr. Seong-Jae Jung, Mr. Dong-Hyun Kim, Mr. Hyung-Sik Kim, Ms. Jeong-Soon Kim, Mr. Kwang-Hyun Kim, Ms. Kyung-Ah Kim, Ms. Kyung-Hee Kim, Mr. Bong-Ki Koh, Mr. Dong-Min Lee, Mr. Hee-Seok Lee, Ms. Jin-Ah Lee, Ms. Kyung-Soon Oh, Mr. Se-Hyun Park, Mr. Chang-Bum Shin, Ms. Soo-Jung Youn, Ms. Seol-Hye Won, Mr. Jung-Hee Yi.

References

1. Nielsen, J.: Return on Investment for Usability, Alertbox (January 7, 2003) Available online at http://www.useit.com/alertbox/20030107.html
2. Marcus, A.: Return on Investment for Usable UI Design, User Experience Magazine(Winter 2002) Available online at http://www.upassoc.org/usability_resources/usability_in_the_real_world/roi_of_usability.html
3. Usability First: Usability ROI: Case Studies. Available online at http://www.usabilityfirst.com/roi/studies.txl
4. Nielson, J.: Corporate Usability Maturity. Alertbox (April 24, 2006) Available online at www.useit.com/alertbox/maturity.html
5. Karat, C.-M.: Business Case Approach to Usability Cost Justification. In: Bias, R.G., Mayhew, D.J. (eds.) Cost-Justifying Usability, Acaedemic Press, Boston (1994)
6. Mayhew, D.J.: Promoting, Establishing and Institutionalizing Usability Engineering. CHI 2004 Full Day Tutorial (April 4, 2003)
7. Rubin, J.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. John Wiley and Sons Inc., New York (1994)
8. Rust, R.T., Thompson, D.V., Hamilton, R.W.: Defeating Feature Fatigue, Harvard Business Review (February 2006)
9. Gussow, D.: Unraveling technology: Manufacturers worry that if a product is complicated, consumers won't buy it. St. Petersburg Times (April 11, 2005) Available online at http://www.sptimes.com/2005/04/11/Technology/Unraveling_technology.shtml
10. Knowledge@Wharton: The Upgraded Digital Divide: Are We Developing New Technologies Faster than Consumers Can Use Them? (October 05, 2005) Available online at http://knowledge.wharton.upenn.edu/article.cfm?articleid=1292&CFID=1763301&CFTOKEN=36902501
11. Bias, R.G., Mayhew, D.J. (eds.): Cost-Justifying Usability. Morgan Kaufmann, San Francisco (1994)

12. Trenner, L., Bawa, J.: The Politics of Usability: A Practical Guide to Designing Usable Systems in Industry. Springer-Verlag, Heidelberg (1998)
13. Lee, D.-S., Woods, D.D., Kidwell, D.: Escaping the design traps of creeping featurism: Introducing a fitness management strategy. Usability Professionals' Association Conference (UPA 2006), Broomfield, CO, USA (June 13–16, 2006)
14. Norman, D.: The Design of Everyday Things, Basic Books (1988)
15. Norman, D.: Simplicity Is Highly Overrated, jnd.org (2007) Available online at http://www.jnd.org/dn.mss/simplicity_is_highly.html

An Axiomatic Method for Cross Cultural Usability Analysis

Sheau-Farn Max Liang

Department of Industrial Engineering & Management, National Taipei University of Technology, 1, Sec. 3, Chung-Hsiao E. Rd., Taipei 10608, Taiwan, ROC
maxliang@ntut.edu.tw

Abstract. Cross cultural influences on usability should be investigated together with human cognition and perception, and the context of use. In practice, to reveal culture similarities is more important than differences. An axiomatic method for cross cultural usability analysis was proposed for tackling these issues. It was argued that usability problems related to human cognition and perception can be identified through the Independence Axiom, whereas the best design can be recognized through the Information Axiom with the domain-specific knowledge.

Keywords: Axiomatic Design, Cross Cultural Usability, Culture Similarities.

1 Introduction

Globalization of markets and applications of information technology have made people around the world able to easily contact with each other and access to a vast of information. As a result, cultural diversity in terms of backgrounds of users and contents of information now becomes more important issue for usability than at any time in the past.

Usability is defined as "the 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 [1]." A product designed for high usability facilitates the completion of relevant tasks in an effective, efficient, error-free, and satisfactory manner. Usability is a key factor for differentiating the products from others. In current merging global markets, thoughtful consideration of regional user needs and desired functionality combined with culturally-sensitive design will greatly increase the chances that we meet our customers' expectations.

Due to its parsimoniousness, the dispositional approach has been widely applied to cross cultural research. Among many studies, the most extensive work has been done by Hofstede [2] on modeling cultural diversity through his five national culture dimensions. A brief description for each dimension is listed below:

High/Low Power Distance: A group in which the degree for the less powerful members to expect and accept that power is distributed unequally is high/low

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 355–364, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Individualism/Collectivism: A group in which the ties between individuals are loose/tight

Masculinity/Femininity: A group in which social gender roles are distinct/overlap

Strong/Weak Uncertainty Avoidance: A group in which the degree for members to feel threatened by uncertain or unknown situations is strong/weak

Long-/Short-term Orientation: A group in which virtues are oriented towards the future/the past and present

Other models with more or less similarity are Trompenaars and Hampden-Tuner's [3] 7D model (i.e., *Universalism/Particularism*, *Individualism/Communitarianism*, *Neutral/Emotional*, *Specific/Diffuse*, *Achievement/Ascription*, *Attitudes to Time*, and *Attitudes to the Environment*) and Victor's [4] LESCANT model (i.e., *Language*, *Environment and Technology*, *Social Organization*, *Contexting*, *Authority Conception*, *Nonverbal Behavior*, and *Temporal Conception*). In addition to these models, Hall [5] observed from different cultural groups and proposed two dimensions of culture differences described as follows:

Monochronic/Polychronic Time: A group in which the things are scheduled one at a time/many in parallel

High/Low Context in Communication: Large/Small amount of stored or unspoken information in a given communication

Research on cross cultural usability has shown that culture differences in term of the index values on these culture dimensions can play an important role in determining the performance of people in many activities, such as the research on air transport crews about their culturally-driven attitudes toward automation and communication [e.g., 6], on jobs of an information system to be matched with users' culturally-specific behavior [e.g., 7], and on international customers' web searching performance [e.g., 8]. Despite its usefulness in capturing the influences of cross cultural differences on usability, this dispositional approach has its drawbacks when applied to the research on cross cultural usability:

- Effects of culture may not sufficient to explain the overall usability of a product. To be more complete and accurate, culture should be studied together with human cognition and perception.
- In practice, to reveal cross cultural similarities is more important than the differences for the design of products. Research should focus more on the similarities than the differences.
- To consider the culture backgrounds of users alone is not sufficient. The context of use should be also taken into account in the research.

Details of these arguments are provided in following sections.

1.1 Study Culture Together with Human Cognition and Perception

From anthropological point of view, culture is defined as the total pattern of human behavior and its products or artifacts that reflect a set of values, norms or standards shared, learned and accepted within a range of variation by members of a particular group of people [9]. Different from human basic physical and psychological

functionality which are inherited and universal to all people, culture is learned and specific to group or category [2]. This distinction corresponds to the difference between the bottom-up processing and the top-down processing in human perception [10]. That is, human cognition and perception are influenced by the human sensory systems as well as the experience and knowledge in mind. A good design has to meet basic ergonomic requirements (e.g., accessibility, free of confusion) first then to consider cultural requirements.

1.2 Focus More on Culture Similarities Than Differences

Culture-related issues are usually addressed through the processes of internationalization and localization. While the process of internationalization is to develop a general or culture-neutral base for localization, the process of localization is to meet the language or cultural requirements from specified target markets or locales [11, 12]. In practice, localization is usually an expensive and inefficient process [13]. Developing culturally neutral products may not only reduce the cost of production but also enhance the share in markets. There has been an attempt in search of internationalized operator interface displays in process control, and resulted a great cross cultural similarities among several Asian culture groups [14].

1.3 Study Culture Together with the Context of Use

The dispositional approach defines culture groups by their most probable values, norms or standards. The variation within a culture is treated as random errors. Fig. 1 demonstrated the distribution for a specific culture group against a culture dimension could approximately be a normal distribution [3].

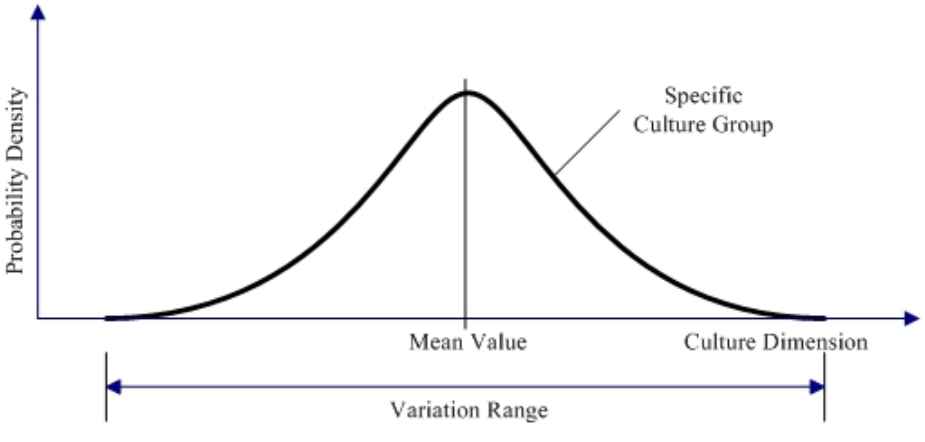


Fig. 1. Distribution for a specific culture group against a culture dimension

In most cross cultural research, within-culture variations are much greater than between culture differences [15]. Thus, to treat the within-culture variations as just errors in the data and ignore them may not be appropriate. Research showed that

significant part of within-culture variations could be explained by the context of use [16, 17]. That is, users from the same culture group may behave differently in different contexts of use. The context of use includes users, equipment, tasks and goals, and environment. In the checklists of ISO 9241-11:- Guidance on Usability [18], social/cultural environment is one of usability criteria in the category of environments, together with the usability criteria in the categories of users, equipment, and tasks. A domain-specific knowledge and a understanding of regional culture are necessary for designs with sound usability.

A new promising usability analysis method is under development [19], and it seems capable to deal with the cross cultural usability issues mentioned above. This method is based on the Axiomatic Design (AD) theory [20]. The AD theory is introduced below followed by a case study to demonstrate its application procedures.

2 Axiomatic Design

Axiomatic Design (AD) approach [21, 22] is a tool for designers to construct and understand design problems, as well as to find possible solutions. AD has been widely applied in the designs of software applications, consumer products, manufacturing systems, and decision support systems [23]. AD views the design process as a series of mappings between four domains: the customer domain, functional domain, physical domain, and process domain. The objective of the AD is to establish a scientific foundation for design activities by two axioms [20]:

Axiom 1: *The Independence Axiom*: Maintain the independence of functional requirements.

Axiom 2: *The Information Axiom*: Minimize the information content in design.

The most applied mappings are the mappings between Functional Requirements (FRs) in the functional domain and Design Parameters (DPs) in the physical domain. The independence axiom claims that each FR should be satisfied by the mappings between FRs and DPs without affecting other FRs, that is, the independence of FRs. Relevant to the information theory [24], the information axiom indicates that the best design is the design with minimum information content. In statistical terms, the best design has a set of DPs to fulfil their associated FRs with the highest probability of success. The mappings between FRs and DPs can be defined as below:

$$\{FR_n\} = [A]_{nm} \bullet \{DP_m\} \tag{1}$$

Where $\{FR_n\}$ is the n-vector of FRs in the functional domain, $\{DP_m\}$ is the m-vector of DPs in the physical domain, and $[A]_{nm}$ is called a design matrix of $\{FR_n\}$ and $\{DP_m\}$. The binary values of elements in the design matrix represent the mapping relationship between $\{FR_n\}$ and $\{DP_m\}$. While the value of 0 denotes no relationship between associated FR and DP, the value of 1 stands for the full relationship between them.

The relationship between a set of FRs and a set of DPs is categorized into three types of design: uncouple design, decouple design, and coupled design. A 3×3 design matrix is used as an example to illustrate these three design types:

$$\begin{Bmatrix} FR_1 \\ FR_2 \\ FR_3 \end{Bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{Bmatrix} DP_1 \\ DP_2 \\ DP_3 \end{Bmatrix} \tag{2}$$

Where FR_1, FR_2, and FR_3 are three FRs in the functional domain. DP_1, DP_2, and DP_3 are three DPs in the physical domain. a_{ij} (i, j = 1, 2, or 3) is the element of the design matrix.

When $a_{ij} = 1$ for all i = j, and $a_{ij} = 0$ otherwise, the design is an uncoupled design illustrated as Eq. 3.

$$\begin{Bmatrix} FR_1 \\ FR_2 \\ FR_3 \end{Bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} DP_1 \\ DP_2 \\ DP_3 \end{Bmatrix} \tag{3}$$

When $a_{ij} = 1$ for all $i \geq j$, and $a_{ij} = 0$ otherwise, the design is a decoupled design illustrated as Eq. 4.

$$\begin{Bmatrix} FR_1 \\ FR_2 \\ FR_3 \end{Bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{Bmatrix} DP_1 \\ DP_2 \\ DP_3 \end{Bmatrix} \tag{4}$$

If a design is neither an uncoupled design nor a decoupled design, then it is a coupled design. An example of a coupled design is illustrated as Eq. 5.

$$\begin{Bmatrix} FR_1 \\ FR_2 \\ FR_3 \end{Bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{Bmatrix} DP_1 \\ DP_2 \\ DP_3 \end{Bmatrix} \tag{5}$$

Note that only the uncoupled design satisfies the Independence Axiom. That is, one-to-one mappings between FRs and DPs.

In the Information Axiom, the information content is measured by its information amount. The information amount is defined as the probability of satisfying a certain FR. For example, if the probability of satisfying the FR_i is P_i, then its information content, I_i, is defined as Eq. 6 as below:

$$I_i = \log_2 \frac{1}{P_i} = -\log_2 P_i \tag{6}$$

From Eq. 6, if $P_i = 1$, then $I_i = 0$, which means the FR_i is satisfied in one hundred percent. When the value of P_i approaches to 0, the value of I_i approaches to infinity, which means the FR_i is almost impossible to be satisfied.

These two design axioms can be applied to the new design of products, manufacturing processes, or systems, as well as to the evaluation and improvement of existed designs. The procedure is first to eliminate any decoupled or coupled design by

applying the Independence Axiom. If there still have more than two alternatives remained, the second step is to select the design with minimum information content by applying the Information Axiom.

3 Case Study: Alarm Icon Design

From a previous study [25], a set of icons used in a DCS product for ASEAN market was reviewed and used as an example to show how the axiomatic method can be applied as a systematic framework for the design of icons in the alarm summary operator interfaces.

An alarm is generated whenever an abnormal condition occurs. By clicking the "Alarm Summary" button on the tool bar of the system home page, the "Alarm Summary" page would be shown on the screen. Typically, 12 alarms could be displayed simultaneously on a single screen. The information for each alarm was listed horizontally on the "Alarm Summary" screen with the associated icon shown on the left side of each alarm. On the "Alarm Summary" screen, there were four icons in the legend to represent four different conditions of alarm:

- Acknowledged and in Alarm
- Unacknowledged and in Alarm
- Unacknowledged and Disabled
- Unacknowledged and Returned to Normal

The first step of applying axiomatic method is to define a set of independent referents, i.e., the FRs. After the analysis on the intended meanings of four alarm conditions, the alarms were classified by two sets of independent FRs. The first set was in terms of Alarm Acknowledgement with two FRs:

FR_1: Acknowledged
FR_2: Unacknowledged

The second set was in terms of Alarm Status with three FRs:

FR_1: In Alarm
FR_2: Disabled
FR_3: Returned to Normal

The second step of applying axiomatic method is to design the icons from a set of visual features, or to review the features used in current design.

The blink was used for the current alarm icon design to distinguish the acknowledged alarms and unacknowledged alarms. However, the distinctiveness among the conditions of "In Alarm", "Disabled" and "Returned to Normal" were not clear. Alarm icons represented these different conditions shared the same color (e.g., red) and shapes (e.g., square and asterisk) that might confuse users. Current design is shown in Table 1.

Table 1. Current design of icons

Intended Meaning	Current Design	Description
Acknowledged & in Alarm		Red asterisk
Unacknowledged & in Alarm	 *blink*	Blinking red asterisk
Unacknowledged & Disabled	 *blink*	Blinking red square with a white dash inside
Unacknowledged & Returned to Normal	 *blink*	Blinking red square with a white asterisk inside

The visual features of current design and their further decompositions are listed as below:

DP_1: Animation
  DP_{11}: No Blink
  DP_{12}: Blink
DP_2: Symbol
  DP_{21}: Asterisk
  DP_{22}: Dash
  DP_{23}: Square
DP_3: Color
  DP_{31}: Red
  DP_{32}: White

Now we could apply the Independence Axiom for the current design:

$$\begin{Bmatrix} \text{Acknowledged} \\ \text{Unacknowledged} \end{Bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} \text{No Blink} \\ \text{Blink} \end{Bmatrix} \tag{7}$$

$$\begin{Bmatrix} \text{In Alarm} \\ \text{Disabled} \\ \text{Return to Normal} \end{Bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{Bmatrix} \text{Asterisk} \\ \text{Dash} \\ \text{Square} \end{Bmatrix} \tag{8}$$

$$\begin{Bmatrix} \text{In Alarm} \\ \text{Disabled} \\ \text{Return to Normal} \end{Bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{Bmatrix} \text{Red} \\ \text{White} \end{Bmatrix} \tag{9}$$

The results showed that only Eq. 7 was uncoupled design. Both Eq. 8 and 9 were coupled design. This analysis revealed that the usability problem of the set of icons was due to the confusion, the failure to discriminate similar stimuli that represent different concepts [10]. It was a universal problem across all culture groups. Once this problem has been solved, the next step is to apply the Information Axiom to find out

the best set of icons. Since users may be interfered or facilitated by their cultural background and the context of use for interpreting the icons, a user survey or testing is an effective tool for designers to gain their domain-specific knowledge. It is suggested here that the information content, I_i, can be measured through the probability of successful associations between to-be-used visual features and their referent concepts. The best design then can be selected according to the Information Axiom. For example, in Fig. 2, through a user survey or testing, there are several options (i.e., Shape 1-5) can be chosen as possible design solutions for representing a referent, but there may be only one option with the largest success percentage of representation as the best choice since it is with the least information content (i.e., Shape 3 in Fig 2). Cross cultural similarities or differences can also be examined via checking how similar or different the profile of the success percentage does by comparison with the one from other culture group.

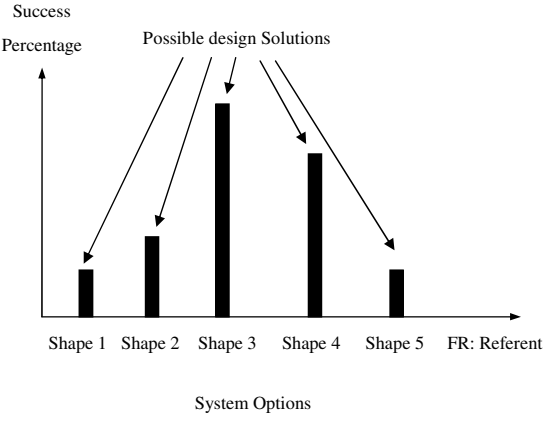


Fig. 2. System options, possible design solutions, and the best solution

4 Conclusion

Cross cultural design is more complex than is always apparent. Variations in the degree to which preferences and values are held within any single culture. Beliefs and values can be dynamic, shifting over time due to social change. Cultures and the context of use frequently interact with unexpected results. All of these factors conspire to make it more difficult to predict consistent behavioral effects of cultures. Most of the research and practice has focused on cross cultural differences and applied a dispositional approach. In contrast, cross cultural similarities have less been researched and the context of use has been less emphasized. The axiomatic method in this paper is a first attempt in this endeavor to highlight the human cognition and perception, and the context of use should be considered together with culture in the cross cultural usability research.

Acknowledgments. The author would like to express his gratitude to the National Science Foundation in Taiwan (NSC 95-2221-E-027-081-MY3) for its support.

References

1. ISO (International Organization for Standardization). ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability. ISO, Geneva (1998)
2. Hofstede, G.: Cultures and Organizations: Software of the Mind. McGraw-Hill, New York (1997)
3. Trompenaars, F., Hampden-Tuner, C.: Riding the Waves of Culture: Understanding Cultural Diversity in Business. Nicholas Brealey, London (1997)
4. Victor, D.: Cross-cultural factors in negotiations: The LESCANT approach for accommodating cultural differences in international legal practice. In: International Practitioners' Workshop Series, vol. XI, pt. 3, American Bar Association, Chicago (1996)
5. Hall, E.T.: The Dance of Life: The Other Dimension of Time, Doubleday, New York (1983)
6. Helmreich, R.L., Merritt, A.C.: Culture at work in Aviation and Medicine: National, Organizational, and Professional Influences, Ashgate, Aldershot, UK (1998)
7. Lee, H.: Time and information technology: Monochronicity, polychronicity and temporal symmetry. European Journal of Information Systems 8, 16–26 (1999)
8. Rau, P.-L P., Liang, S.-F.M.: Internationalization and localization: Evaluating and testing a web site for Asian users. Ergonomics 46(1/3), 255–270 (2003)
9. Haviland, W.A.: Cultural Anthropology, 8th. edn., Harcourt Brace, Fort Worth, TX (1996)
10. Wickens, C.D., Lee, J.D., Liu, Y., Gordon Becker, S.E.: An Introduction of Human Factors Engineering. Pearson Prentice Hall, Upper Saddle River, NJ (2004)
11. Aykin, N. (ed.): Usability and Internationalization of Information Technology. Lawrence Erlbaum Associates, Mahwah, NJ (2005)
12. Fernandes, T.: Global Interface Design: A Guide to Designing International User Interfaces, AP Professional, MA (1995)
13. Prabhu, G.V., Chen, B., Bubie, W., Koch, C.: Internationalization and localization for cultural diversity. Proceedings of the 7th International Conference on Human-Computer Interaction 1, 149–152 (1997)
14. Liang, S.-F.M., Khalid, H.M., Taha, Z., Plocher, T.: In Search of Internationalized Operator Interface Displays in Process Control: A Comparison among Malaysian, Singaporean and Chinese. In: Proceedings of the 7th International Conference on Work With Computing Systems, pp. 253–258 (2004)
15. Shweder, R.A., Sullivan, M.A.: The semiotic subject of cultural psychology. In: Pervin, L.A. (ed.) Handbook of Personality: Theory and Research, Guilford, New York, pp. 399–416 (1990)
16. Hong, Y., Chiu, C.: Toward a paradigm shift: From cross-cultural differences in social cognition to social-cognitive mediation of cultural differences. Social Cognition 19(3), 181–196 (2001)
17. Briley, D.A., Morris, M.W., Simonson, I.: Reasons as carriers of culture: Dynamic versus dispositional models of cultural influence on decision making. Journal of Consumer Research 27, 157–178 (2000)
18. Smith, W.J.: ISO and ANSI Ergonomic Standards for Computer Products: A Guide to Implementation and Compliance. Prentice Hall PTR, NJ (1996)

19. Lo, S., Helander, M.G.: Developing a formal usability analysis method for consumer products. In: Proceedings of the Third International Conference on Axiomatic Design, ICAD-2004-26, 8 pages (2004)
20. Suh, N.P.: The Principles of Design. Oxford University Press, New York (1990)
21. Suh, N.P.: Development of the science base for the manufacturing field through the axiomatic approach. Robotics and Computer Integrated Manufacturing 1(3/4), 399–455 (1984)
22. Suh, N.P., Bell, A.C., Gossard, D.C.: On an axiomatic approach to manufacturing systems. Journal of Engineering for Industry, Transactions of American Society of Mechanical Engineers 100(2), 127–130 (1978)
23. Suh, N.P.: Axiomatic Design: Advances and Applications. Oxford University Press, New York (2001)
24. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana, IL (1949)
25. Liang, S.-F.M., Plocher, T.A., Lau, W.C., Chia, Y.T.B., Rafi, N., Tan, T.H.R.: Usability evaluation on Honeywell process control workstations. In: Proceedings of the 5th Annual International Conference on Industrial Engineering- Theory, Applications and Practice, ID 202 (CD-ROM) 6 pages (2000)

The Impact of Culture on Usability: Designing Usable Products for the International User

Carol Lodge

Human Interactive Technologies, Inc.
Jamaica, West Indies
clodge@gmail.com

Abstract. The purpose of this paper is to examine the impact of culture on the usability and design of global applications. Specifically, this paper will serve to address the theoretical implications of Hofstede's cultural dimensions and the impact of these cultural models on designing usable global products. The paper concludes with a discussion regarding best practices for designing international products.

Keywords: Culture, usability, International User.

1 Introduction

The impact of culture in designing global applications such as websites and cell phones can be regarded as one of the most overlooked aspects of the technology product development cycle as companies try to save costs by developing a generic to serve all users. Frequently, companies are inundated with the marketing vision of a product without considering the importance of integrating culture in the development and use of global applications.

What is culture? How does culture impact the development and design of a global application? Culture by definition is often reported within the Human Computer Interaction body of literature as having values and behaviours shared by a group of individuals. Each culture can have its own values, behaviours which may be defined by certain elements such as language, colours, symbols, or icons. The influence of childhood, education and society can also affect the way individuals interact with other groups. Each particular culture will share similar attitudes as well as think and act similarly in certain situations. Cultures can also be defined by boundaries or regions within a country where various languages are spoken. A culture grouped according to cultural conventions may also exist within an organization or company. For example, employees in a company may belong to a culture as they share a common bond or group that resides in their company.

With the recent emergence of companies marketing products globally, the increasing interest in culture by manufacturers and developers is evidenced by the fact that culture and its impact on usability is an important factor in the product development process which directly influences international users who use these applications.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 365–368, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Depending on the cultural environment of a user, the user may focus on aspects concerning the usability and satisfaction of global applications. As a result, it is important that user interface design elements such as date and time formats, navigational structures, colours, symbols, icons and instructional direction are addressed before global applications are marketed internationally.

2 Cultural Dimensions and Usable Product Design

One of the most often quoted and influential researchers within the field of cross-cultural research is Geert Hofstede [2], who published a landmark cultural study of hundreds of IBM employees in 53 countries over a six-year period, also confirms how cultures perceive things differently. In his cross-cultural studies, Hofstede analyzed how patterns of acting, feeling and thinking are often ingrained in people by late childhood, and any differences in these cultural patterns are displayed in the choice of symbols, rituals and values by a culture. He created five dimensions of culture and the 53 countries surveyed were rated for each dimension on a scale of 1 to 100. These dimensions of culture included power-distance, individualism/collectivism, masculinity/femininity, uncertainty avoidance and short-term/long-term orientation.

1. Power Distance dimension involves the extent to which less powerful members expect and accept unequal distribution of power within a society. This dimension is measured by the Power Distance Index (PDI). Cultures with a high PDI tend to be attracted to leaders with a certain level of dictatorship, and teachers as well as parents are highly respected. Conversely, cultures with low PDI expect more equality, supervisors and subordinates are closer, and organisations have flatter structures.
2. Individualism vs. Collectivism dimension refers to cultures with loose ties where everyone is expected to look after themselves and their immediate family. Collectivism believes in strong group relationship where loyalty is dominant.
3. Masculinity vs. Femininity dimension refers to gender roles within a culture. Masculine cultures exhibit qualities that are assertive, tough, and competitive while feminine cultures exhibit family oriented and tenderness roles.
4. Uncertainty Avoidance dimension involves cultures that behave differently with regards to uncertainty. Cultures with low uncertainty avoidance are more comfortable with uncertainty. On the other hand, cultures with high uncertainty avoidance prefer rules, can be resistant to change and uncertainty may result in high anxiety.
5. Short-term vs. Long-term Orientation dimension examines the degree to which a culture embraces traditional values to new age values or way of life.

Relative to Hofstede's cultural dimensions Marcus and Gould [3] outlined how each of these dimensions can influence aspects of user interface and web design. These are as follows:

1. Power Distance – Information access, mental models hierarchies, value given to authoritative and official symbols.
2. Individualism vs. Collectivism - Interfaces may reflect importance given to personal achievement, sense of social morality and emphasis on change.
3. Masculinity vs. Femininity – Masculine cultures would focus on user interfaces and web design that offer such elements as gender/family/age traditional values,

and user navigation emphasizing exploration and control, whereas the blurring of gender roles and mutual cooperation would be important interface elements for feminine cultures.

4. Uncertainty Avoidance – High uncertainty avoidance cultures will tend to prefer such elements as using clear metaphors, navigational structures that prevents the user from being lost. Interface elements for low uncertainty avoidance cultures will include such elements as less controlling use of navigation and the use of color, typography to emphasize information.
5. Short-term vs. Long-term Orientation – Long-term cultures would emphasize aspects of user interface design to include elements such as the use of relationships as a basis for credibility and practice and practical value focused content. Likewise, interface elements for short-term cultures would focus on content based on truth and the use of rules as a basis for information and credibility.

Similarly, cultural markers have been used to facilitate culturability which is a term coined by Barber et al [1] used to describe the relationship between culture and usability. They proposed the development of cultural markers as elements prevalent or preferred in a particular culture should be identified and incorporated within web design. Examples of cultural markers are national symbols, colours, icons, fonts and belief systems that contribute to the design of web systems that directly affect the user's interaction with the interface.

The development of global applications should accommodate the user's cultural background and environment. Before global products are designed, companies will need to ensure that the correct information is implemented by validating the above design and influences with the users in their target cultures.

3 Four Strategies for Designing Usable International Products

In marketing products internationally other strategies that would prove useful and should be considered when developing usable global applications are as follows:

3.1 Know the International User

When designing international usable products it is important to assess the user's culture, education and behaviours. As in the development of any usable product the first question should ask what users hope to achieve through the use of this product. A detailed design document should be maintained that ensures all design goals are clearly explained about user requirements, etc.

3.2 Know the User's Language

English is considered the official language of only a few countries, which is a small percent of the world's population. However, most websites and applications are in English. While many users speak English as a second language, most prefer to converse in their native language. As companies continue to conduct business globally, the need to localize applications becomes more important. For example, localising a website into languages other than English can result in the site reaching a larger part

of the global online population. However, deciding to localise an entire site depends on the viability of international market opportunities and can lead to several challenges. Before localizing a site companies should consider languages that are appropriate for the target market. For example, include multiple languages within one country. For example, consider incorporating French and English for Canadian websites. Consideration should also be given to user elements such as dates, time, etc. Different countries have different conventions for the use of date and time. For example, the date format used for the US is month/day/year, while other countries use day/month/year. Always use words for the months and use a four digit number for the year.

3.3 Know Cultural Implications and Use of Colour

The integration of colour into the user interface and web design may influence the user's expectations about navigation and content. It is important to assess the use of colour and it's meaning within certain cultures as the use of colours maybe perceived as being negative in one culture and positive in another. The recruitment of qualified international users and usability specialists to test the appropriate use of colour in designing international websites can prevent disasters from occurring.

3.4 Know Cultural Use of Symbols

Symbols or icons used on international user interfaces can convey various meanings for different cultures. Designers should avoid references and symbols that will not be properly translated from one culture to the other.

4 Conclusion

To ensure that the final design is globally usable, professionals and users within your target countries should be used to evaluate your application before going public as they can reveal problems that were not taken into account during the design phase. Additionally, the application or website should be tested regularly for improvements. As the usage of international applications increase, the challenge of enabling more people from various countries to use content and tools effectively will depend increasingly upon global usability solutions and cultural understandings. By attending to the needs of International users with the development of usable applications and websites, companies will achieve greater success and increased profitability.

References

1. Barber, W., Badre, A.: Culturability: The Merging of Culture and Usability. Human Factors and the Web (1998)
2. Hofstede, G.: Cultures and Organizations: Software of the Mind. McGraw Hill, New York, NY (1997)
3. Marcus, A., Gould, E.: Cultural Dimensions and Global Web User-Interface Design. Interactions 7(4), 32–46 (2000)

A Digital Training System for Freehand Sketch Practice

Ding-Bang Luh[1] and Shao-Nung Chen[2]

[1] Institute of Creative Industry Design
Department of Industrial Design
National Cheng Kung University
luhdb@mail.ncku.edu.tw
[2] Department of Industrial Design
National Cheng Kung University
soloncku@yahoo.com.tw

Abstract. Freehand sketch is a fast and easy tool for idea development and communication, especially in the critical front-end or predevelopment stages. While it is important to any designer, lacking of appropriate mechanism for correction in fundamental design education makes even professionals hard to precisely handle accuracy of perspective sketch. Based on two-point perspective method and using cubic shape as subject, this research develops reverse drawing approach and accordingly establishes a digital training system for freehand sketch practice, namely Perspective Practice. Users can operate conventional pen and paper for input and the system automatically illustrates on screen a correct perspective drawing on top of the sketch done by the user, pointing out the concept or technique for improvement. The system provides users with ways in understanding their current skills and guidelines for improvement, through which the efficacy in digital technical training can be enhanced.

Keywords: freehand sketch, perspective drawing, reverse drawing method, digital training.

1 Background and Goal

Freehand sketch is a basic skill to any industrial designer and a fast and easy tool or media for idea development and communication. Through sketches or 3D perspective, designers are able to interact with inspiration, to record, compare, organize, and improve their ideas towards new product concepts. Through sketches, designers are able to discuss with clients and exchange ideas with engineers more effectively and efficiently to improve product design and to facilitate new product development process. With swift advancement of computer technology, many conventional tools for creative development and communication have been replaced by system software. This is particularly obvious in downstream processes when concerned information is clearly defined. In front-end development stages, characterized by high degree of

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 369–378, 2007.
© Springer-Verlag Berlin Heidelberg 2007

uncertainty, computer system is hard to do anything and sketching remains a key competence.

Sketch ability is viewed as one key factor for recruiting designers by almost all design managers. There is a strong link between one's sketch ability and one's creativity or imagination of the future. While important, sketch training has been lacking of quality control mechanism in fundamental design education. Because of that, many professional designers remain suffering from poor perspective sketches, including perspective errors, inconsistence between perspective sketching and engineering drawing, askew horizon line, to name a few. Major causes are three:

1. High teaching load in related courses. Averagely, one teacher instructs 30 to 40 design students in presentation techniques or perspective drawing class. With very limited time for each student, individual errors are hard to be completely corrected.
2. Students are trained to undertake creative works in the future, and so they are often required to practice work without firm answer. When a student has problems in perspective sketching, the problems can hardly be identified and corrected by one self, nor can they be instructed by peers, senior students, or even the teacher.
3. Perspective sketch bases on visual recognition, not on mathematical precision. Thus, it is difficult to develop objective measures for improvement.

Among the above causes, if the third one can be effectively managed, then the rest can be eased accordingly. With an emphasis on the third cause, this research develops needy theories and system tools for digital training and measures for correction, which may reduce reliance of teacher's instruction effort on the student's basic practice while increasing opportunity for self-training, enhancing efficacy in digital learning.

Perspective methods can be categorized into three approaches: one-point, two-point, and three-point. Due to various needs of professional work, designers with different backgrounds adopt different approaches. The one-point method is commonly used by interior designers, two-point approach by industrial designers, and three-point process by architects and urban designers. Relatively, two-point method has merits of extensive applicability and being easier to learn, and thus widely adopted by product designers. It is therefore concentrated in this research.

Perspectives deal with cognition of 3D space, in which shapes distort and the quality and quantity of firm information for interpretation are insufficient. Hence, it is hard to establish the intended system in accordance with individual points in the space. Exercise of cubic shape is fundamental to all design students. Based on practice of cubic shape, concerned theories and procedures are developed.

To date, pen and paper are still used as major tools for sketching by most designers for many reasons, high availability, low cost, no need of electricity, among others. The system to be developed- Perspective Practice- employs conventional pen and paper as input device and user interface. Through standard digital equipment Intous 3 by Wacom, perspective sketch is drawn on paper and at the same time automatically input into the system for further computation, comparison and correction. Based on

locations of key points on the sketch and through the reverse drawing method to be introduced in latter section, the system defines all critical information for a correct perspective sketch and redraws a correct one on its top, pointing out error types made by the learner and suggesting useful guidelines for refinement.

2 CAS and CAI Systems

With advancements of CAD, AI, pattern recognition and related technologies, computer aided sketch (CAS) has become a research focus [6]. CAS concerns design behavior, interaction between object in construction and thinking process in mind, among others. Although claimed to put an emphasis on design concerns, it essentially deals with engineering issues to date. Most CAS systems, such as Viking [7], and algorithms employed in CAD systems [8] adopt parallel perspective as analytical element, which is a drawing tool or language for engineers, not for product designers. Most product designers are trained to first draw 3D perspectives to meet customers' needs in visual perception, form recognition and product aesthetics, and then interpret them into engineering drawing for communication with engineers and NPD work partners. The quality of their 3D perspectives largely determines the fate of a new product idea in development. In light of its importance, designer's drawing quality issue is targeted in this research.

Literature survey reveals that existing computer aided instruction (CAI) systems for perspective sketch mainly function to demonstrate construction of perspective drawing. [1], [2], [3], [4]. These systems are essentially electronic version of textbooks, merely instructing users how to construct a 3D perspective. Functions that help students to identify their error types and to correct and enhance their sketch skills have not been developed yet. If such mechanisms can be implemented, students can train and improve all by themselves, which makes digital learning on technical skills more effective and efficient.

Since there is no closely related system for reference, this research intends to develop one that bases on two-point drawing method for training basic 3D perspective sketch skill, using cubic shape as exercise subject.

3 Foot-Point Method and Its Required Information

There are two ways in construction of two-point perspective drawing: foot-point and measure-point approaches. In this paper, the former method is adopted. For successful exercise, the following information is required (Fig. 1): from top view, picture line (PL) location (for easy drawing, PL is aligned with front corner of the square in order to have an actual length for use later on), side length (L) and rotated angle (R) of the square, and station point (S) location, i.e., the length between PL and drawer or viewer (segment length of SA or SB); from side view, horizon line (HL) location, foot-point (FP) location, and length between object and viewer's eye (E) or foot

(in this case the cube is placed on the ground). Since foot-point approach is fundamental to students of engineering colleges and design schools, detailed drawing procedures are not introduced herein.

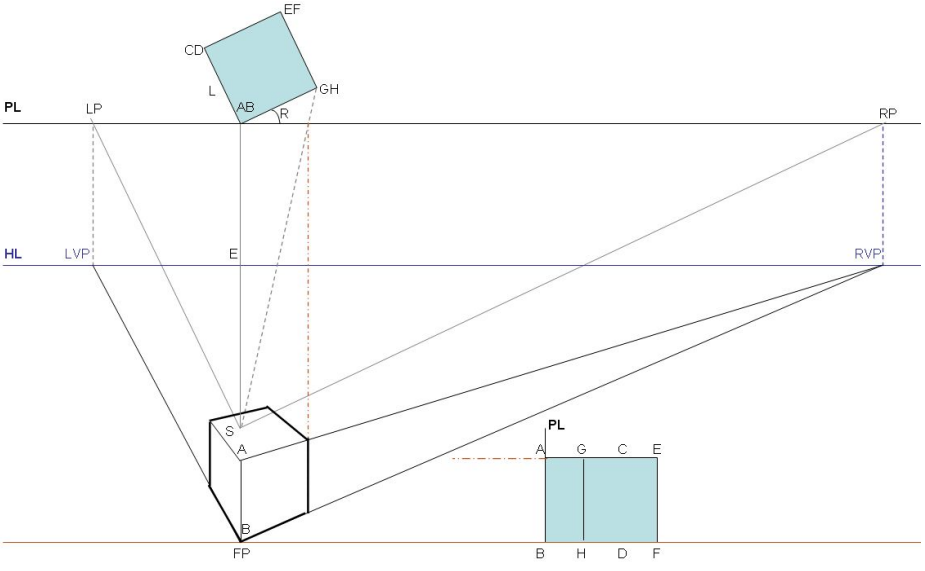


Fig. 1. Foot-Point Approach and its Required Information

Error types in perspective sketching of cubic shapes are (Table 1): vanishing lines in parallel, excessive vanishing points, anti-perspective, proportion maladjustment, askew center line of vision, askew horizon line, and beyond cone of vision [5]. Methods for detecting perspective errors or guidelines for improvement include (Table 2): slope analysis, minimal angle, inclusive circle, and back side comparison. Based on the above information, the system checks on perspective quality and identifies error types and provides guidelines for improvement. Precision is hard to achieve in freehand sketch. Tolerance should thus be a design factor. In accordance with visual capability, a threshold of ±2° is applied to any line drawing for viable perspective sketch of cubic shape.

4 Perspective Practice System

Following due perspective drawing methods, a viable 3D cubic shape can be obtained in accordance with its engineering drawing. By reversing the drawing process, key elements for correction of a cubic shape sketch can be defined. Accordingly, error types of the sketch and suggestions for improvement can be made.

The new system consists of two components- basic exercise and advance refinement- of which details are elaborated as follows:

Table 1. Error Types in 3D Perspective drawing

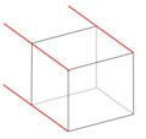
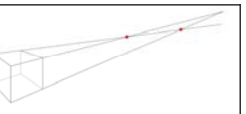
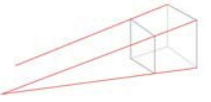
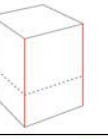
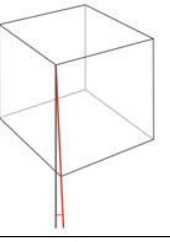
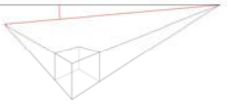
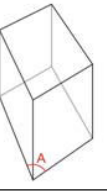
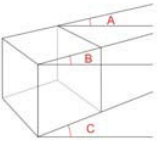
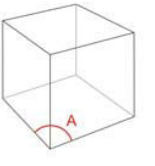
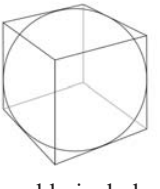
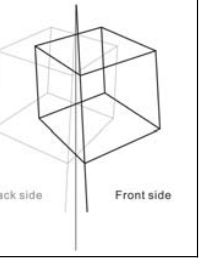
(1) vanishing lines in parallel		(5) excessive vanishing points	
(2) anti-perspective		(6) proportion maladjustment	
(3) askew center line of vision		(7) askew horizon line	
(4) beyond cone of vision			

Table 2. Methods for Perspective Check

(1) slope analysis	 A< B< C	(3) minimal angle	 90°< A
(2) inclusive circle	 Roughly include a circle	(4) back side comparison	

4.1 Basic Exercise Component

This component allows user to choose a cubic shape in space from selected angles (in top view, rotation angles from 15° to 75° with an increment of 15°; in side view, from -30° to +30° due to constraint of the cone of vision) for exercising perspective sketching (Fig. 2). Through the pen-paper-Intous 3 interface, the sketch and the chosen shape are overlapped for comparison and correction. The main functions are three folds: (1) to train the user to memorize certain cubic perspectives for easy application, (2) to let the user understand the error types one often makes, and (3) to enhance the user's freehand technique to an acceptable level for further refinement.

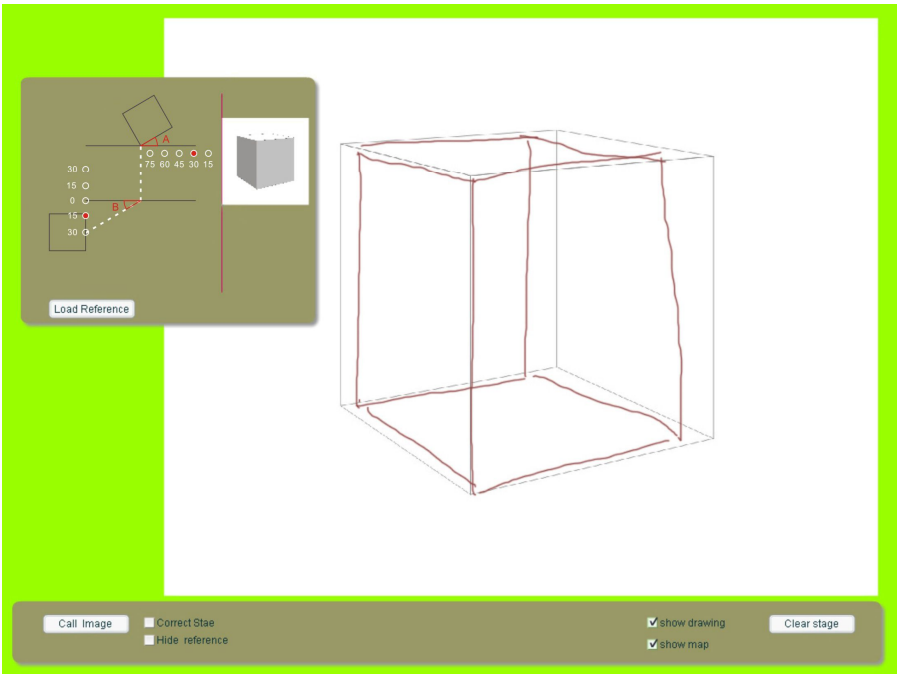


Fig. 2. Illustration of the Basic Exercise Component

Based on the following rules, the system identifies and shows the error types that users make:

1. If any vertical line is tilted more than ±2°, then the system shows the vertical line in red and display "askew vertical line."
2. If length of the central vertical line is shorter than any of the vertical lines on two sides, then the system displays "anti-perspective."
3. If length of the central vertical line is equal to any of the vertical lines on two sides, then the system displays "vanishing lines in parallel."
4. If horizon line is tilted more than ±2°, then the system displays "askew horizon line."

5. If included angle between the two lines on the top or at the bottom is less than 90°, then the system displays "beyond cone of vision."
6. In each set of perspective lines in parallel, have the one in the middle as center line and the corner point where the center line passes through as base point, if the center line intersects with the other two lines at different points and the ratio of the length from the base point to the first intersection to that to the second intersection is less than 0.85, then the system displays "excessive vanishing points"
7. If length of the central vertical line is greater then the distance between those on the sides, then the system displays "thin cube" or "proportion maladjustment."
8. If length of the central vertical line is shorter than any other non-vertical line, then the system displays "flat cube" or "proportion maladjustment."

4.2 Advanced Refinement Component

Through the basic exercise component, a viable cubic shape is ready for further refinement. Based on 6 reference points (a, b, c, d, g, h) on the sketch cube (Figure 5), the system defines center of vision (CV) line (line ab), datum line (either line bd or line bh) and datum plane (left side or right side of the cube) for interpretation of a correct answer. A qualified datum plane must meet the following requirements: it should be the one that with shorter length from CV line to the vertical line of its far side, and that its upper and lower perspective lines form an acceptable vanishing point. When one side plane is defined as datum plane, the lower perspective line of the other side plane is defined as datum line. (Fig. 3)

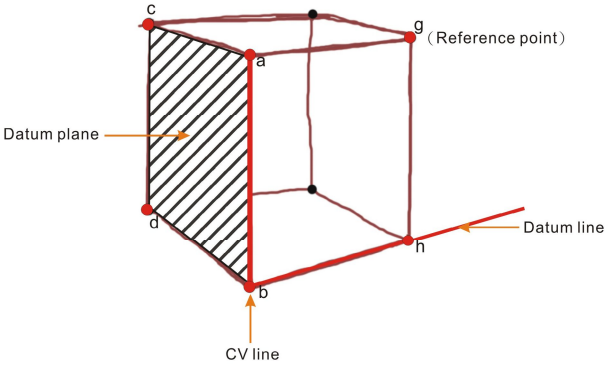


Fig. 3. Reference Points, CV line, and Datum Line and Plane for Advanced Refinement

In accordance with the above drawing information and based on the following "reverse" drawing process (Fig. 4), the system redraws a correct cube superimposing up on the sketch cube for comparison and further refinement of one's sketch skill. (Fig. 5)

1. Have CV line ab as vertical axis and get one vanishing point (LVP) from the datum plane. From point LVP, draw horizon line HL which is perpendicular to CV line, by extending datum line bh begets the vanishing point on the other side (RVP).
2. The cone of vision can be defined by drawing a circle at the center point (o) of the vanishing points with a radius of half of the length between two vanishing points. The station point (s) can be identified by extending CV line to the circle of the cone of vision.
3. Provided that there exists a picture line (PL) which is in parallel to horizon line HL. The CV line extension intersects with the picture line PL at point x. Respectively parallel to line sRVP and line sLVP, draw two line segments xy and xw, of which lengths are identical to that of line ab.
4. Extend line cd to intersect with PL at point j, draw line wj to intersect line ab at standing point (SP)
5. Line SPy intersects with picture line PL at point k. The vertical line of point k intersects with line bh at point h', where is the correct location of point h. Follow the same method to define the correct location of point g.
6. At intersect of line cRVP and line g'LVP is the correct location of point e.

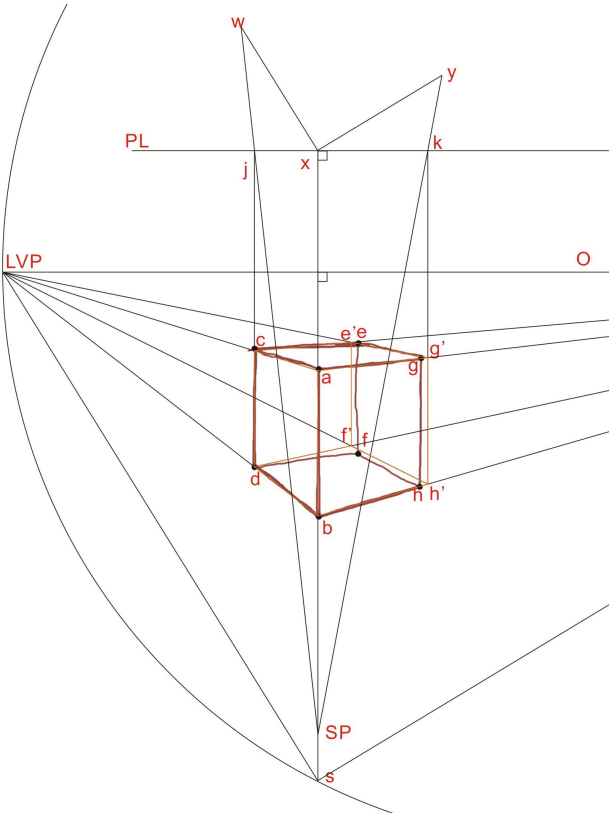


Fig. 4. Reverse Drawing Process for Perspective Construction

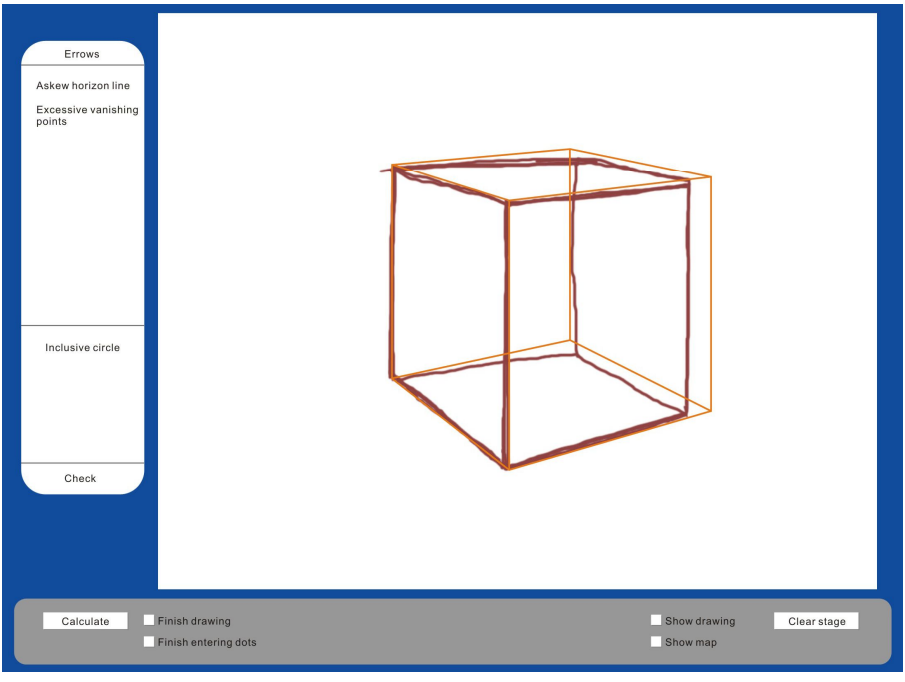


Fig. 5. Illustration of the Advanced refinement component

5 Conclusions and Suggestions

Based on visual recognition and the reverse perspective drawing process, the Perspective Practice system is developed for digital training of perspective sketching. It consists of two components, one for enhancing basic skill and another for further refinement. With this system, students have right answers for exercise and their basic skills can be enhanced efficiently and effectively. Current accomplishments are as follows:

1. The system is practical and useful for digital training of basic perspective skill.
2. With this system, users can examine the error type most frequently made and accordingly improve their concepts and techniques.
3. The reverse perspective drawing process is feasible and innovative.
4. In perspective space, each straight line may have its unique slope, through which the spatial relations among different lines can be clearly defined. This might provide a new solution to help ease the face recognition difficulty in computer aided sketch.

The usability and effectiveness of the new system will be tested in Presentation Techniques class, Spring semester of 2007, by students of Department of Industrial Design, National Cheng Kung University, Taiwan. Concerned statistical results will be presented at the conference.

References

1. Leung, C.Y., Wu, C.T.: The Research and Design of Computer Assisted Instruction, Tatung University, Taipei (1998)
2. Hong, W.S.: Software Design of Computer Assister Instruction for Perspective Drawing and Evaluation of its Effectiveness, Tatung University, Taipei (2000)
3. Lin, S.T.: The Design and Research of Computer Assisted Instruction for Light-Shadow Perspective Drawing, Tatung University, Taipei (2004)
4. Lin, C.H.: The Effectiveness and Design of Computer-Assisted Instruction for Descriptive Geometry, Tatung University, Taipei (2004)
5. Luh, D.B., Yang, T.L.: Idea Presentation Techniques, 2nd edn., pp. 74–75. Chuan-Hua Publishing, Taipei (2006)
6. Sun, S.C., Sun, L.Y.: Status Quo and Perspective of Computer Aided Sketch Development. China Mechanical Engineering 17(20), 2187–2192 (2006)
7. Wallace, D., Jakiela, M.J.: Automated Product Concept Design: Univying Aesthetics and Engineering. IEEE Computer Graphics and Application 13(4), 66–75 (1993)
8. Liu, J.Z., Lee, Y.T.: A Graph-Based Method for Face Identification from a Sigle 2D Line Drawing. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10), 1106–1119 (2001)

Culture Issues in Traffic Sign Usability

Annie W.Y. Ng and Alan H.S. Chan

Department of Manufacturing Engineering and Engineering Management
City University of Hong Kong
Kowloon Tong
alan.chan@cityu.edu.hk

Abstract. Traffic signs are probably the best known graphical symbols that we encounter along roads and highways daily in a traffic system. The authors had conducted two experiments with two different groups of Hong Kong Chinese subjects to investigate the usability of traffic signs with guessing and comprehension tasks. The first experiment used Mainland China traffic signs, while the second experiment employed Hong Kong traffic signs. In this paper, the effects of two user factors (Mainland China visit experience and non-local driving experience) and one sign feature (concreteness) on task performance were investigated for exploring the culture issues in traffic sign usability. It was shown that Mainland China visit experience of subjects was a significant factor in affecting their sign guessing performance. The result also indicated that when a specific cultural issue is incorporated in a traffic sign, the sign should be accompanied by supplementary text to reduce the effect of cultural bias. It was interesting to note that non-local driving experience had a negative effect on local sign comprehension when signs were pictorially similar but different in intended messages; but the effect was positive when the signs look alike and conveyed the same meaning. A recommendation to ensure sign comprehensibility for non-local drivers is that a leaflet containing sign information should be made available for vehicle drivers at passport control points. On sign feature, concrete signs that bear a resemblance to actual objects contribute to higher guessability scores than abstract ones, which may be due to the fact that the thinking style of Chinese people is synthetic, concrete, and relies on the periphery of the visible world. Therefore, concrete signs are better than abstract signs in regard of providing visualization aids in helping Chinese subjects to complete the guessing task. The findings revealed the importance of taking the cultural issue into consideration when developing traffic signs, and provided information and recommendations for the design of highly comprehensible traffic signs.

Keywords: culture, usability, traffic sign, sign concreteness.

1 Introduction

1.1 Usability

The International Organization for Standardization [1] defined usability as ‘an extent to which a product can be used by specified users to achieve specified goals with

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 379–387, 2007.
© Springer-Verlag Berlin Heidelberg 2007

effectiveness, efficiency, and satisfaction in a specified context of use'. Jordan et al. [2] identified three distinct components, namely, guessability, learnability, and experienced user performance (EUP) that would influence the usability of a product, and later added system potential and re-usability as further possible components of usability [3]. As suggested by Jordan [4], guessability is measured by the cost (e.g. time and error) in using a product to perform a task for the first time – the lower the cost, the higher the guessability. The learnability component is concerned with the cost to the user in achieving some competent levels of performance on a task with a product, while the component of EUP is the relatively stable level of performance that an experienced product user reaches. The system potential component represents the theoretical optimal performance obtainable from a product with respect to a particular task, and the re-usability refers to the level of performance achieved when a user returns to a task with a product after an extended period of non-use.

1.2 Culture

User factors such as experience, domain knowledge, cultural background, disability, age, and gender were found to have effects on usability [4]. The United Nations Educational, Scientific and Cultural Organization [5] described culture as follows: 'culture should be regarded as the set of distinctive spiritual, material, intellectual and emotional features of a society or a social group, and that it encompasses, in addition to art and literature, lifestyle, ways of living together, value systems, traditions and beliefs.' According to Smith [6], culture is 'the actual practices and customs, languages, beliefs, forms of representation, and a system of formal and informal rules that tell people how to behave most of the time.' In the consideration of conceptual compatibility for design of human machine interfaces, a study on colour associations showed that Hong Kong Chinese, Korean, and Thai did not generally share common colour-concept associations [7]. In the design of a traffic system, an uncontrolled junction for European drivers means an obligation to yield to vehicles on the right, while for American drivers it represents priority for them [8]. On the design of traffic signs, a recent review revealed that rectangular traffic signs give direction or information in China, Japan, France, Spain, Germany, Singapore, Austria, Hong Kong, and Taiwan, while they give notices of requirements, prohibitions or restrictions in America and New Zealand [9].

1.3 Traffic Sign

Traffic signs are used to regulate, warn, and guide road users in a traffic system and probably they are the best known graphical symbols that we encounter along the roads and highways. Even though traffic signs are not standardized across places, many people licensed in one place will drive in other places without any further training or testing. In Hong Kong, visitors (those arrive at Hong Kong other than taking up residence for a period of exceeding 12 months) are allowed to drive if they hold recognized overseas driving licenses or international driving permits [10]. A few cross-cultural studies on comprehensibility of traffic signs have recently been reported ([11], [12]). Hawkins et al. [11] evaluated the comprehensibility of 21 regulatory, warning, and truck-related traffic signs used in the United States for 759

Mexican business and tourist drivers coming from Mexico. The results demonstrated that seven signs including Yield sign, Fasten Safety Belts sign, Right Lane Ends sign, Load Zoned Bridge sign, Weigh Station sign, Hazardous Cargo Route sign, and Hazardous Cargo Prohibited sign had correct plus partially correct response rate lower than 75% (Fig. 1), and 14 signs had comprehensibility higher than 75%. The findings indicated that most of the Mexican drivers had some degree of understanding of the United States traffic signs. Shinar et al. [12] examined the comprehension levels of 31 highway traffic signs for 1,000 licensed drivers from Poland, Canada, Israel, and Finland. Other than fifteen signs commonly used in all countries, two signs were common to two or three countries and fourteen signs were unique to individual countries only. Their results showed that the comprehension levels of locally used signs (78%) were much higher than that of non-local signs (32%).



Yield | Fasten Safety Belts | Right Lane Ends | Load Zoned Bridge | Weigh Station | Hazardous Cargo Route | Hazardous Cargo Prohibited

Fig. 1. Traffic signs had correct plus partially correct response rate lower than 75% (Hawkins et al. [11])

1.4 Aim

As cross-border motoring of goods vehicles, passenger vehicles, and private cars between Hong Kong and Mainland China is steadily increasing these years [13], there is a need to highlight the importance of developing traffic signs with culture issues in mind. In this paper, the effects of two-user factors, viz., Mainland China visit experience and non-local driving experience, and the sign feature of concreteness on guessing / comprehension performance were examined for ascertaining the effect of culture issues on traffic sign usability. It was hypothesized that concrete signs would lead to higher guessability scores than abstract ones. Better guessing ability on Mainland China traffic signs was expected from subjects with more frequent Mainland China visit experience, and non-local driving experience would have a negative effect on the comprehensibility of local traffic signs.

2 Experiments

The authors conducted two experiments with two different groups of Hong Kong Chinese subjects to investigate the usability of traffic signs with guessing and comprehension tasks. The first experiment used Mainland China traffic signs, while the second experiment employed Hong Kong traffic signs.

2.1 Experiment 1

This experiment was done for examining the effects of Mainland China visit experience and concreteness on the guessability of traffic signs with prospective-users.

Nineteen male and twenty-two female Hong Kong Chinese engineering undergraduates who had never taken any driving tests, nor possessed a driving license of any sort, voluntarily participated in this experiment. Their ages were between 18 and 27 years (median = 22.5 years). All subjects had no colour deficiency and reported having no previous experience of learning the meanings of any Mainland China traffic signs.

To minimize the influence of daily encounters or prior experience with traffic signs on the results, instead of using the signs currently used in Hong Kong, 120 traffic signs stipulated in the latest National Standards of the People's Republic of China for Road Traffic Signs and Markings (GB5768-1999 issued in April 1999) were employed. These 120 traffic signs were chosen based on two criteria: firstly, their messages are conveyed with symbols only; secondly, they are not used in accompaniment with other signs for transmitting a message.

Subjects were first required to guess the intended meaning of each sign within 10 seconds, and then to report their Mainland China visit experience in past 12 months. Finally, the subjects were asked to give a subjective rating for the degree of concreteness (0 = definitely abstract, 100 = definitely concrete) for each sign. Traffic signs are regarded as concrete if they depict real objects, materials, or people while those do not are considered as abstract.

2.2 Experiment 2

Experiment 2 addressed the effect of non-local driving experience on the comprehensibility of traffic signs for licensed drivers. One hundred male and nine female Hong Kong full driving license holders were invited to participate in the questionnaire survey. Their ages were between 18 and 57 years (median = 32.5 years). The two criteria used in Experiment 1 were also adopted for choosing signs in this study. There are 178 signs contained in Hong Kong Schedule 1 (Traffic Signs) of the Chapter 374G (Road Traffic Regulations) of Legislation Database and 82 of them satisfy the two characteristics for selection. Twenty one of the 82 signs were randomly selected for testing. Subjects reported their non-local driving experience first, and then completed multiple-choice questions for evaluating their traffic sign comprehension. An interviewer guided each respondent through the questionnaire, if necessary.

3 Results and Discussion

3.1 Experiment 1

Amongst the 120 Mainland China traffic signs, 26 of them were structurally similar to and had the same intended meaning as those in Hong Kong. Subject's guessing performance for similar signs (61.54%) was significantly better than that for dissimilar signs (52.94%) (Paired-samples t test, $p < 0.05$), showing the benefits of using common or similar signs in different regions.

The similar signs with minimum and maximum guessability scores are shown in Table 1 and Table 2. Though this was a guessing task, 16 of the 26 similar signs had guessability scores higher than 50%. The guessing score for sign M5 (keep right) was the lowest, 2.44%. It was wrongly interpreted as turn right, steep descend, and U-turn.

Table 1. The similar signs with minimum guessability scores

	Mainland China	Hong Kong	Meaning	Guessability score		The three most frequent responses
				Mean (%)	Standard deviation (%)	
The lowest six	 M5		Keep right	2.44	15.62	Turn right (27%) Steep descend (10%) U-turn (7%)
	 W16		Road narrows on right	14.63	23.03	Narrow road (22%) Passing allowed (10%) Straight road and curve road (10%)
	 W4		T-junction ahead	17.07	30.84	No-through road (20%) Stop (12%) T-junction ahead (10%)
	 W15		Road narrows on both sides	17.07	24.00	Narrow road (29%) Two roads ahead (15%) Merging into traffic (10%)
	 W26		Slippery surface	36.59	48.77	Slippery surface (24%) Steep descend (12%) Uneven road surface (12%)
	 W28		Riverside way	39.02	20.95	Caution! Fall into the sea (41%) Sea ahead (12%) No entry (5%)

Table 2. The similar signs with maximum guessability scores

	Mainland China	Hong Kong	Meaning	Guessability score		The three most frequent responses
				Mean (%)	Standard deviation (%)	
The highest six	 P20		No left turn	100	0	No left turn (100%)
	 M1		Go straight only	98.78	7.81	Go straight only (88%) Straight road (5%) Upward (2%)
	 P19		No pedestrians	98.78	7.81	No pedestrians (98%) No entry (2%)
	 W23		Mind the signal	97.56	15.62	Mind the signal (98%) No-through road (2%)
	 M2		Left turn only	96.34	17.29	Left turn only (90%) Go straight and then left turn (5%) Turn around (2%)
	 W14		Road works ahead	96.34	17.29	Road works ahead (49%) Work-in-progress (34%) Mining ahead (2%)

3.1.1 Mainland China Visit Experience

Most of the participants (90.24%) visited Mainland China in past 12 months and the remaining 9.76% did not visit there in the same period. The results of Kruskal Wallis test showed a significant effect of Mainland China visit experience on sign guessing ($\chi^2 = 8.554$, df = 1, $p < 0.005$). Those subjects who visited Mainland China during last 12 months could perform better in guessing the meanings of traffic signs (55.62%) than those who did not (47.29%). The result showed that cultural exposure from visits helped subjects interpret the tested signs better and improved guessing performance. It had been shown that the usability of an application across international markets can be increased by the use of icons which are designed to be culture independent [14] or multi-cultural compatible [15]. del Galdo [14] stated that an icon of a black cat has different meanings in different cultures. It is considered good luck in the United Kingdom but bad luck in the United States. The results of this experiment showed that when culture specific graphics are incorporated in a traffic sign, supplementary text may be needed for avoiding cultural bias.

Understanding other people's languages, cultures, societies, political systems, and markets are of great value to travellers. Some signs tested here such as farm trailer prohibited (P10) and tricycle for goods prohibited (P15) were highly related to the Mainland Chinese cultural environment (Fig. 2). Both the farm trailer and tricycle for goods can hardly be found in Hong Kong nowadays. The guessability scores for signs P10 and P15 were 47.56% and 48.78%, respectively. Both scores were substantially smaller than the overall average for all the 120 signs (54.81%). For sign P10 (farm trailer prohibited), only 4.88% of the subjects made a correct response and other subjects guessed it to be freight vehicle, dumper truck, container vehicle, towing vehicle, motorcycle, travelling coach, or maintenance truck prohibited. None of the subjects correctly interpreted the sign P15 as tricycle for goods prohibited.

Although these two signs are somewhat culture dependent and contain specific culture features, their guessability scores were not the lowest ones, indicating that factors other than the cultural issues such as sign concreteness also influenced the guessing performance.



P10 - Farm trailer prohibited



P15 - Tricycle for goods prohibited

Fig. 2. Examples of test signs that were highly related to the Mainland Chinese cultural environment

3.1.2 Concreteness

As expected, concrete signs contributed to higher guessability scores than abstract ones (r = 0.387, n = 120, $p < 0.001$). Concrete signs bear a resemblance to actual objects while abstract signs do not. The thinking style of Chinese people is synthetic, concrete, and relies on the periphery of the visible world [16]. Therefore, a concrete sign provides a better visualization aid than an abstract one in helping the Chinese subjects to complete the guessing task. Where signs are abstract, access to meaning is much more difficult. This was shown to be the case in the most abstract sign M13

(traffic has priority on the main route; ). It had a concreteness rating of 18.32 and a guessing score of 0%. All the participants incorrectly guessed the sign to be, for example, rocket, gate, and go ahead. It is however, necessary to note that the thinking style of some populations, say Americans, tends to be analytic, abstract, imaginative, and linear [16] and different results may have been obtained if such ethic groups were tested. Further research efforts should be devoted to investigate the relationship between sign concreteness and sign guessability for non-Chinese people.

3.2 Experiment 2

Amongst the 109 subjects, only 30.28% of the participants drove outside Hong Kong (e.g. Mainland China, the United Kingdom, Australia, Canada, Thailand, Taiwan, Macau, America, and Japan) and the remaining 69.72% did not. Analysis of variance showed that non-local driving experience had an effect on local sign comprehension ($p < 0.05$). The average comprehension performance for licenced holders with non-local driving experience (67.96%) was worse than those without non-local driving experience (70.83%), which may be due to retroactive interference in sign information. It was suggested that retroactive interference occurs when new and somewhat similar information disrupts retrieval of information learned earlier [17].

Some traffic signs used in one place can be found in other places, but the intended messages are different. For example, a sign which consists of a heavy red circular border and a white background () depicts 'all vehicles prohibited' in Hong Kong but 'all vehicles and pedestrians prohibited' in Mainland China. Sign I2 () represented a 200-meter distance to an exit from a road, however, a similar sign used in Canada () denoted hazard close to the edge of the road [18]. The comprehension score of sign I2 for drivers with non-local driving experience (61.11%) was lower than that for drivers without the experience (72.37%). Sign W9 () depicted pedestrians on or crossing road ahead, but a similar sign used in Mainland China () represented pedestrians only [19]. The comprehension score of sign W9 for drivers with non-local driving experience (25.93%) was slightly lower than that for drivers without the experience (27.63%). Hence, it is not surprising that non-local driving experience had a negative effect on the performance of sign comprehension. Nevertheless, the results indicated that there is a reinforcement effect with the signs which look alike and have similar meanings. For sign W4 (level crossing with barrier ahead; ), the comprehension score for drivers with non-local driving experience (15.74%) was significantly greater ($p < .05$, Wilcoxon test) than that for drivers without the experience (3.95%). These two signs can be found in other areas like the United Kingdom [20], Macau [21], Mainland China [19], and Taiwan [22].

Testing traffic signs with drivers from different countries (Poland, Canada, Israel, Finland), Shinar et al. [12] found that the comprehension levels for locally used signs were much higher than that for non-local signs. Nevertheless, the findings here demonstrated that while non-local driving experience had a positive effect on local sign comprehension when the signs looked alike and conveyed the same meaning, it reduced the comprehension of the local signs which were similar but different to the intended messages of the signs used in other places.

As cross-border motoring is steadily increasing [13], a recommendation to ensure sign comprehensibility for non-local drivers is that a leaflet with information of traffic signs and corresponding meanings should be made available at control points and handed out to vehicle drivers at passport control. In future, symbolic traffic signs should be standardized across different regions and countries as much as possible, be easy to guess, or have their meanings spelled out in text.

4 Conclusion

This paper was successful in examining few cultural issues in traffic sign usability with guessing and comprehension tasks. Visiting experience of a place was shown to influence sign guessing performance. Non-local driving experience was shown to exhibit different effects on the comprehension of different signs. Concrete signs that bear a close resemblance to actual objects contributed to higher guessability scores than abstract ones. It was believed that different results may have been obtained if non-Chinese subjects were tested. The results here revealed the importance of developing traffic signs with cultural considerations, and provided useful recommendations for the design of traffic signs for culturally diverse users.

References

1. International Organization for Standardization.: 9241-11 Ergonomics Requirements for Office Work with Visual Display Terminals (VDTs) Part 11 Guidance on Usability (1998)
2. Jordan, P.W., Draper, S.W., MacFarlane, K.K., McNulty, S.-A.: Guessability, Learnability, and Experienced User Performance. In: Diaper, D., Hammond, N. (eds.) People and Computers VI. Cambridge, Cambridge University Press on behalf of the British Computer Society, pp. 237–245 (1991)
3. Jordan, P.W.: What is Usability? In: Robertson, S. (ed.) Contemporary Ergonomics 1994, pp. 454–458. Taylor & Francis, London (1994)
4. Jordan, P.W.: An Introduction to Usability, pp. 8–16. Taylor & Francis, London, Bristol, PA (1998)
5. United Nations Educational, Scientific and Cultural Organization.: Universal Declaration on Cultural Diversity (2002)
   http://www.unesco.org/education/imld_2002/unversal_decla.shtml
6. Smith, K.: Handbook of Visual Communication - Theory, Methods, and Media, p. 523. L. Erlbaum, Mahwah, NJ (2005)
7. Chan, A.H.S., Han, S.H., Nanthavanij, S.: Color associations for Hong Kong Chinese, Korean, and Thai - A Comparison. In: Proceedings of IEA 14th Triennial Congress, Seoul Korea (2003)
8. Summala, H.: American Drivers in Europe - Different Signing Policy may cause Safety Problems at Uncontrolled Intersections. Accident Analysis and Prevention 30, 285–289 (1998)
9. Xia, C.: Traffic Sign World. 1st edn. Beijing Shi, Zhongguo Ji Hua Chu Ban She, 3 (2002)
10. Hong Kong Transport Department.: How to Apply for a Driving License (2003) http://www.td.gov.hk/public_services/licences_and_permits/vehicle_and_driving_licences/how_to_apply_for_a_driving_licence/index.htm

11. Hawkins Jr., H.G., Picha, D.L., Lopez, C.A.: Mexican Driver Comprehension of U.S. Traffic Control Devices. Transportation Research Record 1628, 15–24 (1998)
12. Shinar, D., Dewar, R.E., Summala, H., Zakowska, L.: Traffic Sign Symbol Comprehension - A Cross-Cultural Study. Ergonomics 46, 1549–1565 (2003)
13. Hong Kong Customs and Excise Department.: Cross-Boundary Vehicle Movements(2005) http://www.customs.gov.hk/eng/statistics_cross_e.html
14. del Galdo, E.: Internationalization and Translation - Some Guidelines for the Design of Human-Computer Interfaces. In: Nielsen, J. (ed.) Designing User Interfaces for International Use, pp. 1–10. Elsevier, Amsterdam (1990)
15. Goonetilleke, R.S., Shih, H.M., On, H.K., Fritsch, J.: Effects of Training and Representational Characteristics in Icon Design. International Journal of Human-Computer Studies 55, 741–760 (2001)
16. Yang, K.S.: Chinese Personality and its Change. In: Bond, M.H. (ed.) The Psychology of the Chinese People, pp. 106–170. Oxford University Press, Hong Kong (1986)
17. Henderson, J.: Memory and Forgetting, pp. 76–82. Routledge, London, New York (1999)
18. Ontario Ministry of Transportation.: Road Signs in Ontario (2005) http://www.mto.gov.on.ca/english/traveller/signs/
19. Yang, J.L., Liu, H.X.: GB 5768-1999 Dao Lu Jiao Tong Biao Shi He Biao Xian Ying Yong Zhi Nan. Zhongguo Biao Zhun Chu Ban She, Xin Hua Chu Ban She (1999)
20. United Kingdom Department for Transport.: The Highway Code (2004) http://www.highwaycode.gov.uk/signs_index.shtml
21. Government Printing Bureau.: Macau (DL) 17/93/M (1993) http:// www.imprensa.macau.gov.mo/bo/i/93/17/declei17_cn.asp
22. Taiwan Area National Freeway Bureau.: Traffic Signs and Markings (2003) http:// www.freeway.gov.tw/en_07.asp

International Remote Usability Evaluation: The Bliss of Not Being There

Mika P. Nieminen, Petri Mannonen, and Johanna Viitanen

Helsinki University of Technology, Department of Computer Science and Engineering,
P.O. Box 9210, 02015 TKK, Finland
{mika.nieminen,petri.mannonen,johanna.viitanen}@tkk.fi

Abstract. This paper describes the planning and implementation of a cross-border usability test that was to be executed in five European countries. The usability evaluation was designed by the Usability Group at Helsinki University of Technology who also performed the testing for the Finnish partner. In the other countries the usability tests were to be implemented by teams of subject matter specialists with very heterogeneous disciplines ranging from software engineering to social sciences, gender equality and vocational counselling. This paper describes the level of materials and training prepared for the remote usability testing and discusses its adequacy both via test personnel satisfaction and comments, and by comparing the found usability problems and observed phenomena in the test sessions between the test executed by the usability experts and the subject matter specialists.

Keywords: International usability testing, remote usability testing, localizing usability test materials, usability testing by non-expert evaluators.

1 Introduction

Usability testing has proven its worth as a crucial part of software engineering. Faster and wider communication mediums have made distribution of knowledge work, both geographically and temporary, an everyday practice. The change has created networked product development teams/communities and international cross-country organizations. Also usability engineering must be able to perform in these distributed surroundings. Most commonly applied method of usability engineering is conducting usability tests. It is often sighted that evaluating the usability of a product is very sensitive to the social, lingual or contextual environment where the testing is done. In many cases this has lead to increased costs, when international product testing must be executed in several countries by locally hired usability experts. The obvious alternative to boost the efficiency of international or multi-site usability testing has been to develop methods and procedures to conduct the usability tests remotely [1,2,3].

This paper describes the remote usability evaluation of an Internet portal for vocational counselling. The case project, funded by the European Commission's education and culture programme, produced a dynamic web site, which was to be

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 388 – 397, 2007.
© Springer-Verlag Berlin Heidelberg 2007

localized to each of the five partner countries. The development project had identified the need for good usability and its development tasks included usability evaluation of the portal's first functional version. Due to the fact that the portal was to have five different language versions with local content, the usability testing had to be performed locally at each partner country. The challenge was that only one of the project partners had usability expertise at their disposal. This lead to a situation where the tests were planned and coordinated by the members of the Usability Group at Helsinki University of Technology, while half of the actual user test sessions where executed in a distributed fashion by the local subject matter specialists.

In the literature remote usability testing is categorized based whether it happens in real time (synchronous or asynchronous) [3] and further whether the test data is collected automatically or presented by the users themselves [2]. Synchronous remote testing refers to monitoring the test via video connection to the test site or when possible by sharing the to-be-tested application via a broadband network connection using collaboration suites such as Microsoft Netmeeting™, Lotus Sametime™ or ShowMe™ from Sun Microsystems. The above classifications are combined in Fig. 1 with some examples of evaluation tools [4,5,6,7,8].

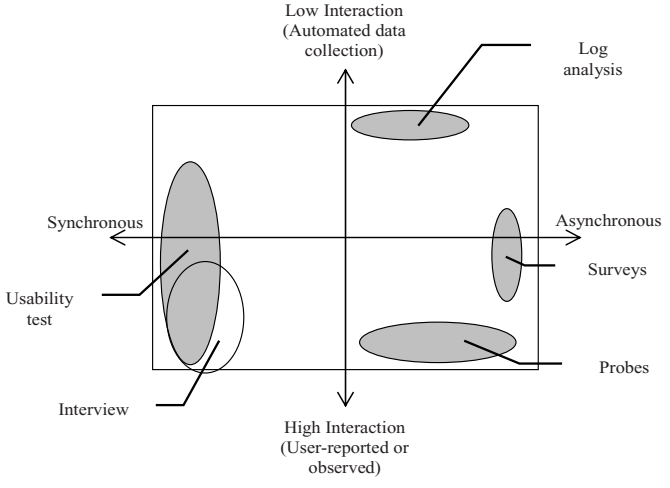


Fig. 1. Categorization of User – Evaluator interaction in remote usability evaluation activities

In this paper we describe an additional category of remote usability testing that is not only distributed geographically, but is also asynchronous. In our variation some of the user test sessions where organized and moderated by persons not fluent in usability testing based on a tailored test manual and a few hours of usability training. Our hypothesis was that with well-targeted instructions and minimal training the local personnel can manage the test sessions and with enough accuracy report the critical event they observe during the tests as shown by experiments conducted by Castillo et al [9,10]. Also by using the local subject matter experts (i.e. people fluent with the project's domain instead of usability engineering) our goal was to distill the cultural

anomalies that would have been incorporated to the analysis if we would have observed the user tests via a translator.

2 Evaluation Procedure

As stated earlier the evaluated system was an Internet portal for vocational counseling. Thus, the identified user groups were job-seekers (primary target group) and vocational counselors (secondary target group). Even thou, none of the project partners were native English speakers it was used as official language and the product development was done in English. During the development the portal was to be localized from the English development version into five different European languages: German, Danish, Slovenian, Romanian and Finnish.

The overall planning of the used usability evaluation methodology and the extent of the user tests were done in co-operation with the Austrian project coordinator and the members of the Usability Group at Helsinki University of Technology (later in this paper referred as "we"). Three user groups to be involved in testing were selected to be: low skilled job-seekers, medium skilled job-seekers and (high-skilled) counselors.

All together 17 users were to be involved in the distributed user tests. Table 1 illustrates the project partners and their planned number of test users.

Table 1. Breakdown of the participants to be included to the user tests

Target groups **Partners**	Low skilled	Medium skilled	Counselors
Project coordinator (Austria)			2
Partner 1 (Austria)	2		
Partner 2 (Romania)			2
Partner 3 (Denmark)			2
Partner 4 (Slovenia)	2		1
Partner 5 (Finland)	2	2	2

The complete usability evaluation procedure was to include the following seven stages:

1. Expert evaluations in Finland
2. Planning of the usability tests
3. Usability tests in Finland
4. Preparation of the test manual, including additional guidelines and checklists
5. Training sessions for the non-expert evaluators about practicalities of usability testing

6. Remote usability test sessions by the non-expert evaluators
7. Aggregation and analysis of the overall results.

Each of these stages is described in more details in the following chapters.

2.1 Expert Evaluations

In the first stage the expert evaluations were performed in Finland during June 2006. We performed expert evaluations for the English version of the portal using heuristic usability evaluation method and Nielsen's 10 heuristics [6]. Due to delays in the development process, some parts of the portals were missing and it could not be evaluated in its entirety. The result of the expert evaluation was a list of 94 prioritized usability problems including six catastrophe and 18 major problems. Usability problems were rated on a scale 1-4 with 4 being the most severe [6]. Conducting an expert evaluation in the beginning of the evaluation procedure allowed us to familiarize ourselves with the portal and find out its overall usability before planning the user tests.

2.2 Planning of the Usability Tests

After the expert evaluations we were able to plan the user tests. The user test were planned to be as simple as possible to help the remote test moderators to run the test sessions as easily as possible. Traditional usability test [6,11], using thinking aloud technique was chosen and its physical requirements were kept to a minimum. The test environment only needed a computer with Internet connection and a video camera to record observations for later analysis.

Two to three evaluators were to be present at each test session; a moderator was responsible for running the test session and other evaluator(s) were responsible for taking notes and observing the test situation. The test setting was explicitly designed not to require a fully furnished usability laboratory with a one-way mirror and multiple video recorders.

Test sessions were planned to last 45-90 minutes and consist of the following parts:

1. *An introduction*, during which the moderator briefly explains the user the test procedure and other ethical consideration [6] and asks the user to complete the pre-test questionnaire. The pre-test questionnaire requests background information such as personal details (age, sex, education and job description or study subject), and asks user open ended questions about her use of information technology and experiences with Internet services.
2. *The actual test tasks*. During the test tasks the user is asked to think aloud while performing the tasks. The user is given, one at a time, 13 tasks covering the core features of the portal, including seeking job descriptions and specific job details, conducting a skills test and an aptitude test, and searching information about available training from the portal. Five tasks included in the test setup required a modified scenario, only a variation in a few words, to cater to the different user groups; job-seekers and counselors.

3. *The debriefing after the tasks*. At the end a drawing assignment is given to the user and she is asked to draw the structure of the system as she recalls it. In the debriefing the moderator also goes through a prepared list of questions about the portal and asks the user for further comments.

2.3 Usability Tests in Finland

The first usability tests were conducted during October 2006. At that time the portal was still under development and it had not been localized into Finnish. All the test materials were in our native language and also the tests were carried out in Finnish. Altogether seven users were selected matching the characteristics of the target groups. A couple of days before the actual tests a pilot test was performed with one user, leaving six users for the actual tests. Our test sessions did take place in our usability laboratory, but the recording was done using a free standing digital video camera instead of the built-in equipment. In each test session three evaluators were present and sessions were conducted according to the set procedure. Afterwards the observation notes were completed and verified by reviewing the video recordings.

After the user tests the findings of the heuristic evaluation and usability tests were combined to produce a master list of usability problems. From those problems a total of 30 subsets were identified and rated using the same severity scale as earlier (catastrophic, major, minor or cosmetic). As a final result we delivered to the project coordinator a standalone evaluation report, in which the severity rated usability problems and corresponding suggestions for improvements were grouped according to the main parts of the portal.

2.4 Usability Test Manual

We used our tests as a baseline to provide the other partners detailed instructions with an accurate observation framework to be used to report the critical incidents in their tests. As said, for the Finnish tests all the test materials were in Finnish and so had to be translated into English for dissemination to the other project partners.

In addition to the actual tests materials, we also provided the project partners with a detailed test specification. This specification, or a usability test manual, did not only cover the actual test procedure, but also the underlining rationale for the specific test tasks. This additional contextual information was needed for the partners to accurately localize the test tasks and scenarios to their respective languages. So, the manual included detailed instructions how to set up and carry out a usability test for an Internet portal for vocational counseling:

- Resource estimate. The manual described briefly the human resources needed for testing, based on having two test users.
- Guidance for localizing the test instructions and tasks. For the remote evaluations all of the materials were to be localized to German, Danish, Slovenian and Romanian languages to match the local versions of the portal. In the manual we argued the reasons why to localize the materials and instructions and how to do it.
- Setting up a usability test. We gave simplified instruction for the non-expert evaluators about how to carry out usability tests including physical test setup and recording, selecting test users, planning test's timetable, and test procedure

(including five phases: preparation, introduction, during the test tasks, debriefing and after the tests). All the necessary test materials were appended to the test manual: a background questionnaire to be filled in by the user at the beginning of the test, individual test tasks to be handed to the user during the test, scenarios and test tasks to be used by the moderator during the test, the drawing assignment, debriefing interview questions, and a checklist for the moderator of a usability test.
- Performing a pilot test before the actual tests.
- Framework for observing and analyzing the success during the tests, and reporting the test data. The local testers were given an observing guide (an observation form with example data and points of interest about the test tasks), which briefly presented a simple analysis strategy for studying the success in tasks. The observation guide and observation form were to help the remote test personnel in making observation notes during the test and reporting findings to project coordinator.

The hypothesis was that by providing comprehensive enough instructions the reported data from the local partners' tests would be comparable and valid.

2.5 Training Session

In mid-October 2006 we provided the project partners a very concise, about five hour training session or introduction to usability testing. The training session was carried out during a project workshop in Graz, Austria and it emphasized the practical side of usability testing and mainly tried to increase the partners' awareness to usability issues. All project partners responsible for testing were present. The training agenda was based on our usability test manual. We gave the partners very brief examples of expected results and experiences from our already conducted usability tests. The actual report of our test results was not delivered to the project partners prior to their respective test sessions. Thus, their observed phenomena and found usability problems were not influenced by our results.

2.6 Remote Tests in Other Partner Countries

The remote international usability tests with 11 users were to be executed without our participation in five European countries solely based on our usability manual and training. Local test moderators interacted with native participants in their respective native language in the local contexts. They were responsible for implementing and running and recording the tests, and reporting the findings. The few qualifications for the local testers were being native speakers of their local languages, attendance to our training session and fluency in written English.

Usability tests in partner countries were carried out in November 2006. The project coordinator reported us the following about the tests:

- The project coordinator had performed tests with one low skilled job-seeker and two counselors using the German language version of the portal. According to the partner all materials, including the checklists, were translated to German before the tests. Two evaluators were present at each test.

- The second Austrian partner took advantage of the already translated materials provided by the project coordinator and also tested the German language version of the portal with two low skilled users.
- Due to scheduling problems the partners in Romania and Denmark could not conduct usability tests at all.
- The Slovenian version of the portal was not completed in time. Instead the tests in Slovenia were conducted using the English version of the portal with one low skilled job-seeker and one counselor. In consequence of not having a localized version the Slovenian partner reported having had problems related to terminology during the tests.

2.7 The Overall Results

The project coordinator was responsible for collecting all the test data from all the evaluations and then analyzing and aggregating the final results. The overall test data consisted of results and suggestions for improvements provided by us (usability inspection and tests with six users) and test data from the local tests provided by other project partners (usability tests with seven users). These aggregate results are to be made a few months after the writing of this paper.

3 Reliability and Validity of the Test Results

As mentioned in the previous chapters we prepared an observation form for the non-expert evaluators to help them report their findings. The provided forms were filled for every one of the realized 13 test sessions. We have used these observation forms to compare the results from the test sessions executed by both usability experts (us) and persons not fluent in usability testing methodology. While the forms did not give us the full richness as if we had attended the usability tests ourselves, they did mark us the critical incidents and gave a rough picture of the tests in general.

The following Table 2 summarizes the main differences between expert and non-expert evaluators when reporting the test data and interacting with the users during the tests.

Table 2. Differences between expert and non-expert evaluators when reporting the test data and interacting with the users during the tests

Type \ Evaluator	Non-expert Evaluators	Expert Evaluators
Reporting the observations	Reported the exact user behavior as a sequence	Reported the user actions in relation to the overall goal
Reported critical incidents	Reported equally all incidents, emphasis on positive comments	Reported incidents relating to usability problems, emphasis on negative comments
Quality of the reported observations	Heterogeneous between the different partners	Uniform among the usability experts
Interaction with the users during the test	Frequent interaction with the users, several assists during a test	Very minimal interaction with the users, assistance involving foreign terms

As the first line in the Table 2 shows we managed to make more observations about the reason why the users did what they did during the test. For instance when we reported how the users interpreted some element in the user interface the non-expert evaluators only reported that the users had difficulties with the element.

The cause for the difference can be explained by the major difference in the observers' experiences with usability testing i.e. their moderator skills. Other option is the individual differences in the users' ability to think aloud or the non-expert evaluators' inability or reluctance to promote the users' thinking aloud.

All in all the results from both experts and non-experts are very consistent. Our findings (based on 6 user tests) cover almost 90% of all the test observations. Similarly the remote tests reported over 70% of our results. All the critical and major usability problems where reported by both groups, except for those arisen from lack of interaction (see the navigation bar example in the following paragraph). Thus, the results from remote tests validated our findings with very good accuracy. In addition there seems to be only a few culture or language specific usability problems.

The single most interesting difference in the observations was the usability problems relating to the portal's navigation bar (including a navigable bread crumb trail) depicted in Fig. 2. In our evaluations none of the test users grasped the functionality of the navigation bar's bread crumb trail and actually only a few noticed or commented the whole bar at all during the tests. In the other hand majority of the remote test users were reported to use the navigation bar, but it is unclear from the reported incidents whether they navigated thru the actual bread crumb trail.



Fig. 2. The portal's navigation bar, with a bread crumb trail

Based on the usability evaluations the portal's user interface was redesigned. The navigation bar was relocated to the top of the screen and its look and feel was changed dramatically. Another major redesign was the removal of the wizard-like bread crumb trail (both from the navigation bar and the job pages) and the introduction of tabbed browsing to bind each job description into a more concrete unit.

A small online survey was done before (N=13) and after (N=30) the user interface changes. In survey, among other questions, three questions where asked relating to the user interface:

1. How would you rate the design of the site?
   (5=Excellent, 1=poor)
2. How would you rate the clarity of the page structure?
   (5=Excellent, 1=Poor)
3. Was it easy to find the information you were looking for?
   (3=Yes, 2=Half-and-half, 1=No)

The below Fig. 3 shows the averages of the survey answers. Even though the survey reached only a relatively small number of people the changed towards better

(or yes) is clear. Uncannily, the improvement in both the design and the clarity is almost equal.

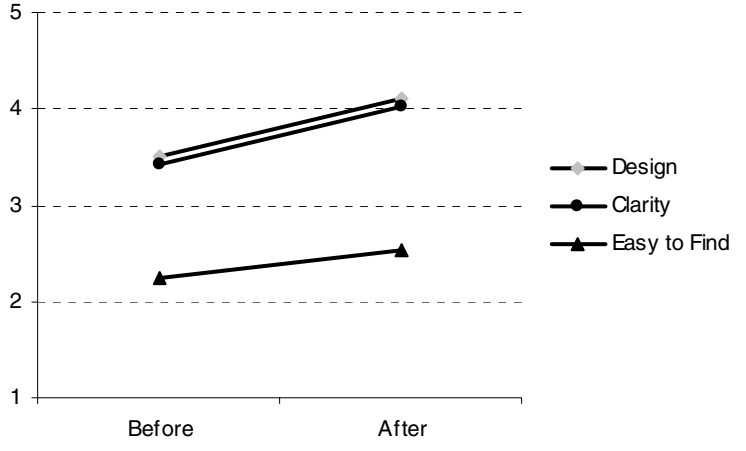


Fig. 3. Before and after user ratings for the portal design, clarity of structure and easiness to find information

4 Conclusions

Remote usability evaluations in an international context can be either very costly or low on results. General guidelines emphasize the risks and obstacles involved in international testing at a distance and guide practitioners towards very traditional, safe and therefore costly usability evaluation methodology [3]. In our study the local subject matter specialists, who were not familiar with usability engineering, were harnessed to execute usability tests in additional two European countries. These non-experts were successful in conducting remote usability tests i.e. recruiting users, organizing and moderating the tests and reporting their observations. The non-expert test moderators perceived the offered materials and training adequate for them to perform the tests. However, our analysis indicates that the observation form provided for the test personnel might have been too suggestive and thus slightly biased the made observations.

Analysis of the reported observations revealed that the results from expert and non-expert tests supported each other almost perfectly. This is in line with earlier studies where minimal training has been reported to give non-experts adequate knowledge to identify, report and rate the severity levels of usability problems they encountered [9,10].

The main difference between the observations of non-experts and experts were the ability to see the big picture (e.g. overall goals) and to produce reports of uniform quality. As suggested by Bojko et al. [12] more test situation training might have helped to make more accurate observation and include also the observers own

interpretation of the user actions. However, it would have also made the process more cumbersome and more expensive.

All in all, the process undeniably produced a better version of the career portal, and the remote test results validate that the portal caters also to the needs of the users in all the partner countries. This shows promise that non-expert personnel can be effectively utilized to carry out usability tests with only minimal training, presupposing there is an experienced usability team coordinating the evaluation.

Acknowledgement. The authors of this paper wish to acknowledge the participation, funding and support of the Leonardo ICT CTO project and the persons therein that conducted the remote usability tests and allowed us to compare their observations to ours.

References

1. Thompson, K.E., Rozanski, E.P., Haake, A.R.: Here, There, Anywhere: Remote Usability Testing that Works. In: Proceedings of the 5th conference on Information technology education (SIGITE'04), pp. 132–137. ACM Press, New York (2004)
2. Krauss, F.S.H.: Methodology for remote usability activities: A case study. IBM Systems Journal 42(4), 582–593 (2003)
3. Dray, S., Siegel, D.: Remote Possibilities? International Usability Testing at a Distance. Interactions Journal 11(2), 10–17 (2004)
4. Gaver, W., Dunne, T., Pacenti, E.: Cultural probes. Interactions 6(1), 21–29, ACM Press, New York (1999)
5. Mattelmäki, T.: Design Probes. University of Art and Design Helsinki, Helsinki (2006)
6. Nielsen, J.: Usability Engineering. Academic Press Inc., New York (1993)
7. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction, 3rd edn. Pearson, London (2004)
8. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-Computer Interaction. Addison-Wesley, New York (1994)
9. Castillo, J.C., Hartson, H.R., Hix, D.: Remote Usability Evaluation: Can Users Report Their Own Critical Incidents? In: Proceedings of the Conference on Human Factors on Computing Sytems (CHI'98): Summary, pp. 253–254. ACM Press, New York (1998)
10. Hartson, H.R., Castillo, J.C.: Remote Evaluation for Post-Deployment Usability Improvement. In: Proceedings of the Conference on Advanced Visual Interfaces (AVI'98), pp. 22–29. ACM Press, New York (1998)
11. Dumas, J.S., Redish, J.C.: A Practical guide to usability testing. Greenwood Publishing Group Inc., USA (1999)
12. Bojko, A., Lew, G.S., Schumacher, R.M.: Overcoming the Challenges of Multinational Testing. vol. 12(6), pp. 28–30 (2005)

A Framework for Evaluating the Usability of Spoken Language Dialog Systems (SLDSs)

Wonkyu Park[1], Sung H. Han[1], Yong S. Park[1], Jungchul Park[1], and Huichul Yang[2]

[1] Department of Industrial and Management En ineerin , POSTECH,
San 31, Hyoja, Pohan , 790-784, South Korea
{p09plus1,shan,drastle,mozart}@postech.ac.kr
[2] Samsung Electronics, Seoul, South Korea
huicul.yang@samsung.com

Abstract. Usability evaluation is now considered an essential procedure in developing a spoken language dialogue system (SLDS). This paper proposes a systematic framework for evaluating the usability of SLDSs. The framework consists of what to evaluate and how to evaluate. What to evaluate includes components, evaluation criteria, and usability measures to evaluate various aspects of SLDSs. With respect to how to evaluate, a procedure for developing scenarios and scenario-based evaluation methods are introduced. In addition, a case study, in which the usability an SLDS was evaluated, was conducted to validate the proposed framework. The results of the case study showed successfully the usability level, usability problems, and design implications for further development. The framework proposed in the study can be practically applied to usability evaluation of SLDSs.

1 Introduction

During the last two decades or so, many studies have been conducted to improve the performance of spoken language dialogue systems (SLDSs). However, most studies focused on recognition performance, while only a few studies investigated human factors issues such as user models, linguistic behavior, user satisfaction, etc. [1]. Human factors issues play an important role in an SLDS because enhanced usability can partially cover imperfect recognition accuracy of the system. When it comes to natural dialogues between the users and an SLDS, it is obvious that value of the SLDS depends on usability, which is critical to make the system commercially successful [2].

Usability evaluation in the development process is essential because it provides current usability levels and reveals potential usability problems. Although a variety of studies conducted usability evaluation [3, 4, 5, 6, 7], only a few proposed systematic evaluation frameworks or methodologies for SLDSs [1, 8]. Walker et al. developed a framework for evaluating SLDSs, PARADISE (Paradigm for Dialogue System Evaluation) [8]. It provides a quantitative usability index (i.e. user satisfaction) considering task success and costs. Dybkjær and Bernsen developed an evaluation template for SLDSs that consisted of 10 entries such as 'what is being evaluated',

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 398 – 404, 2007.
© Springer-Verlag Berlin Heidelberg 2007

'system part evaluated', 'type of evaluation', 'symptoms to look for', etc. [1]. However, these studies are not easy for practitioners to apply to usability evaluation because they do not deal with specific data collection methods.

This paper aims to propose an evaluation framework for SLDSs. The framework identifies usability measures for various aspects of SLDSs. Also, it proposes scenario-based methods to effectively evaluate usability in terms of both performance and satisfaction. In addition, a case study is conducted to validate the proposed framework. An SLDS providing the user information about schedules, contacts, weather, etc. in a home environment, was developed for the case study.

2 Usability Evaluation Framework for SLDSs

The usability evaluation framework proposed in this study consists of what to evaluate and how to evaluate. Fig.1 depicts details of the proposed framework.

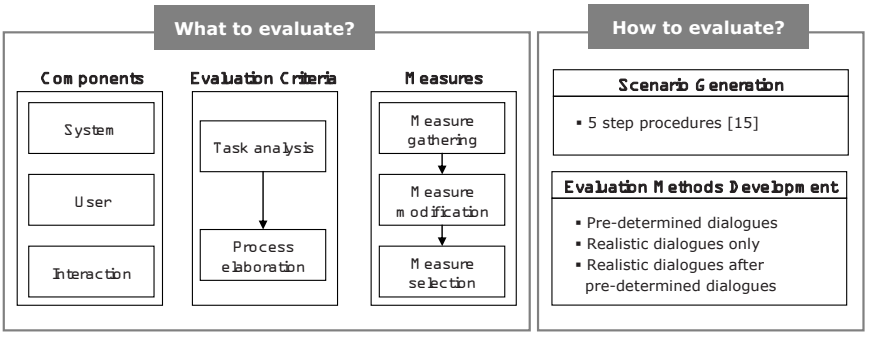


Fig. 1. A framework for evaluating the usability of SLDSs

What to evaluate includes components, evaluation criteria, and usability measures. The framework has three components, i.e. user, system, and the interaction between these two. Evaluation criteria are functions and characteristics of each component that affect the usability of an SLDS. Usability experts identify them using a task analysis technique. Usability of a system can be quantitatively measured by employing performance and satisfaction measures [9]. Relevant measures were surveyed from the existing literatures. From the measures, usability experts selected ones appropriate to each criterion by considering ease of measurement and relevance to usability.

How to evaluate introduces the procedure to create scenarios for SLDSs evaluation and has two methods to collect usability measures. The framework proposes scenario-based evaluation by real users. It enables researchers to find usability problems that originate from mismatches between what the user needs and what the system provides [10, 11].

2.1 'What to Evaluate'

The framework has components, evaluation criteria, and usability measures with respect to what to evaluate. Components in this study are classified into user, system,

and interaction. Various aspects of SLDSs can be evaluated by considering the three components, while previous studies mainly evaluated the system only [1].

Criteria are developed to evaluate each component. The criteria are made by elaborating the process shown in a modified job process chart. A job process chart reported by [12] is a specific type of partitioned operational sequence diagrams. With the modified job process chart, practitioners are able to identify system-user interaction processes and information transmitted between them. An example of the chart is shown in Fig. 2, from which evaluation criteria are made for the case study. For example, an evaluation criterion of 'recognition performance' is elaborated from 'recognize input'. Another example is 'user behavior' that comes from 'construct utterance' when the system fails to provide information that the users request.

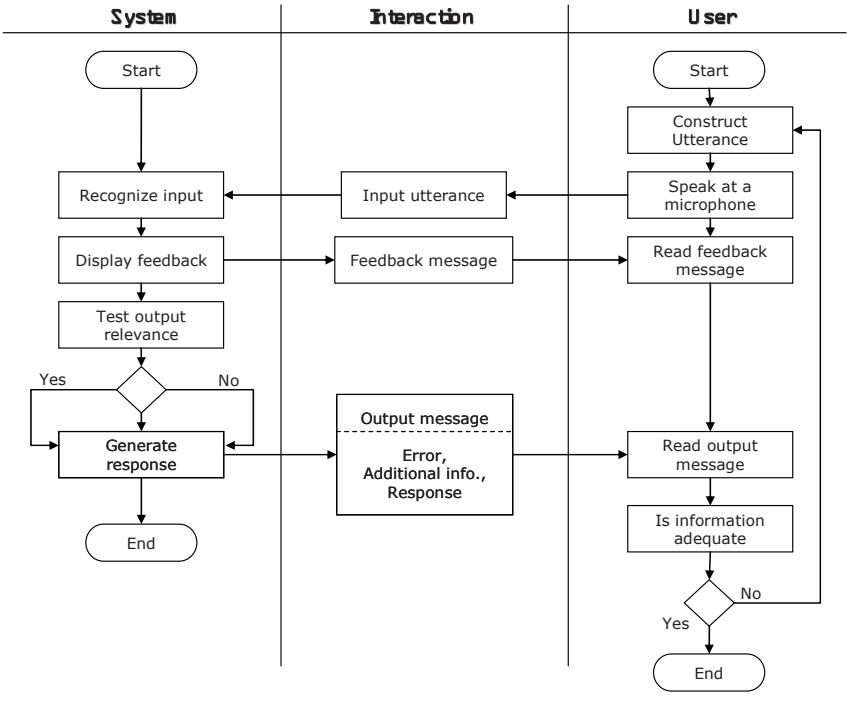


Fig. 2. A modified job process chart of an SLDS used in the case study

A variety of usability measures were collected from previous studies [2, 3, 4, 5, 6, 7]. Some measures (e.g. number of barge-ins and number of SLDS help) were SLDS-specific, while others (e.g. task completion time and number of errors) could be used for general usability evaluations. The latter measures might be modified to fit SLDSs. For example, the number of errors is modified into the number of unrecognized words/utterance and the number of utterance construction errors. Usability measures appropriate to a corresponding criterion were selected by ease of measurement and relevance to usability. For example, word recognition rate was selected to evaluate the

'recognition performance'. Table 1 shows components, evaluation criteria and usability measures developed for the case study.

Table 1. Evaluation criteria and measures of each component for the case study

Components	Criteria	Measures
System	Recognition performance	Sentence recognition rate Word recognition rate Recognition error frequency
	Dialogue model	Adequacy of reasoning function Utterance construction error (frequency and types) Correct response rate
Interaction	System output	User satisfaction on system response
	Task	Task completion time Frequency of failed tasks
User	User satisfaction	Overall user satisfaction on the SLDS
	User behavior	Users' response pattern to various system errors Patterns of utterance construction
	Learning	Utterance variation

2.2 'How to Evaluate'

A scenario-based method can be used to evaluate an SLDS system in a realistic situation [13]. A variety of situations should be considered in evaluation scenarios. However, there exist few studies, except for [14], that systematically develop scenarios reflecting various situations. Park et al. proposed a scenario development procedure that consists of five steps [14]: 1) identifying functions and information that system can provide, 2) analyzing sentence structures appropriate to system functions and information, 3) analyzing proper words for sentence structures, 4) creating scenario structures by mapping words into sentence structures, and 5) developing detail scenarios. This study uses this procedure when creating scenarios.

The framework proposes two scenario-based evaluation methods. The first one uses pre-determined dialogues. It provides utterance that the system can handle. The system can always come up with an answer, unless it fails to recognize the speech pronounced by the user. This method is mainly appropriate to measure recognition performance of an SLDS. Pre-determined dialogues are developed through the entire five steps explained above.

The second method performs scenarios to evaluate an SLDS's overall usability in realistic situations. Given a situation and information to be queried, the user asks the system using his/her own expressions. In addition to recognition performance measured by the first method, the discrepancy between the dialogue model

hypothesized by the developer and the user's actual utterance pattern can be analyzed. Realistic dialogues can be developed using the step 1 stated above. Table 2 shows examples of the two types of dialogues when the user conducts the same task.

The effects of previous experience with the pre-determined dialogues on the user's utterance pattern are also investigated by comparing two user groups (one group of users who conducts the realistic dialogues only, and another group of users who conducts realistic dialogues after experiencing the pre-determined ones).

Table 2. Examples of two dialogue types

Pre-determined dialogues (performed through two transactions)	Realistic dialogues
1: Any e-mail from mom this afternoon?	I heard mom sent me an e-mail this afternoon. So I would like to know contents of the e-mail.
2: Contents of the e-mail?	

3 Validation of the Proposed Framework

A total of 84 subjects who speak Korean participated in the case study. The participants were randomly assigned to one of three different experiments: 60 subjects for conducting pre-determined dialogues (experiment 1), 12 for realistic dialogues (experiment 2), and the other 12 for realistic dialogues after the pre-determined dialogues (experiment 3). A larger number of participants were assigned to experiment 1, because the SLDS was in the early stage of the development process in which the developers needed to focus on the recognition performance. A total of 24 scenarios were developed for the experiments. Twelve scenarios were pre-determined dialogues, while the other were realistic dialogues.

Evaluation criteria and usability measures for the case study were developed according to the proposed framework (See section 2.1), which are shown in Table 1.

The evaluation results provide design problems, usability levels, and valuable design implications for the SLDS. This paper describes sentence recognition rates and correct response rates only. The average values of these measures for the three experiments are depicted in Fig. 3. Based on the results of the usability evaluation, design implications for further development were made. Firstly, the recognition algorithm needs improvement to effectively process users' utterance. The sentence recognition rate of 50 % might be too low for a commercial SLDS. When significant improvement is difficult to achieve, introducing auxiliary input devices such as keyboard and mouse would be a good support for better usability.

Secondly, help documents or training programs should be provided in the SLDS. Experiment 3 that included short training before the main experiment showed better performance than Experiment 2 in both measures. This implies that system help may make it easier for users to use the SLDS. Information describing what and how to interact with the system should be provided for better usability.

Finally, the developers should improve a reasoning function that enables the system to identify user's intention from what it has recognized. It is important when, as in this case, the system's recognition performance is poor. Note that the correct response rates are higher than the sentence recognition rates in all the three experiments. The reasoning function can be improved by refining the dialogue model based on utterance patterns that the users employ in their daily lives.

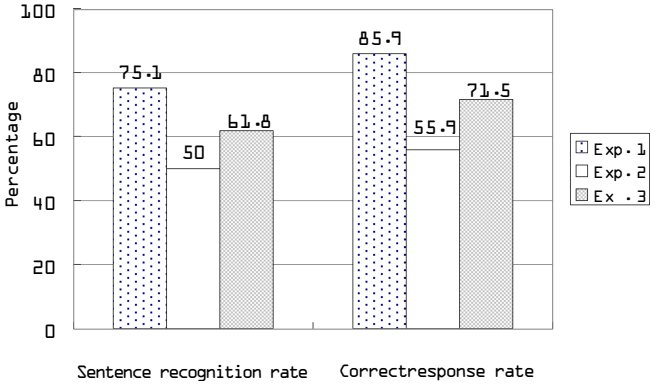


Fig. 3. Correct sentence recognition rates and correct response rates for three experiments

4 Conclusion

A usability evaluation framework for SLDSs was proposed. It focuses on both what to evaluate and how to evaluate. Usability measures are systematically defined to evaluate SLDSs. Evaluation criteria that could affect the usability of SLDSs are identified from a modified job process chart. In addition, the study also proposes two types of scenario-based evaluation methods. Each evaluation method can be used for a different purpose. In a case study, an SLDS was evaluated using the proposed framework. The case study revealed the usability level, usability problems, and design implications for better usability. The framework described in the study can be practically applied to evaluating the usability of SLDSs.

References

1. Dybkjær, L., Bernsen, N.O.: Usability Issues in Spoken Language Dialogue Systems. Natural Language Processing 6, 243–272 (2000)
2. Kwahk, J.: A methodology for evaluating the usability of audiovisual consumer electronic products. Unpublished Ph. D. dissertation, Pohang University of Science and Technology, Pohang, South Korea (1999)
3. Danieli, M., Gerbino, E.: Metrics for evaluating dialogue strategies in a spoken language system. In: The 1995 AAAI spring symposium on empirical methods in discourse interpretation and generation, pp. 34–39 (1995)

4. Dybkjær, L., Bernsen, N.O., Dybkjær, H.: Evaluation of spoken dialogues: user test with a simulated speech recogniser. CPK - Center for PersonKommunikation, Aalborg University 9a & 9b (1996)
5. Litman, D.J., Pan, S.: Designing and Evaluating an Adaptive Spoken Dialogue System. User Modeling and User-Adapted Interaction. 12, 111–137 (2002)
6. Polifroni, J., Hirschman, L., Seneff, S., Zue, V.: Experiments in evaluating interactive spoken language systems. In: The DARPA Speech and Natural Language Workshop, pp. 28–33 (1992)
7. Simpson, A., Fraser, N.A.: Black Box and Glass Box Evaluation of the SUNDIAL System. In: The EUROSPEECH: European Conference on Speech Processing, Berlin, pp. 1423–1426 (1993)
8. Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: PARADISE: A Framework for Evaluating Spoken Dialogue Agents. In: The 35th annual meeting of the association for computational linguistics (ACL-97), Madrid, Spain, pp. 271–280 (1997)
9. 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)
10. Dybkjær, L., Bernsen, N.O.: Usability evaluation in spoken language dialogue systems. In: The Proceedings of the workshop on evaluation for language and dialogue systems, Toulouse, France (2001)
11. Park, Y.S., Han, S.H., Yang, H., Park, W.: Usability evaluation of conversational interface using scenario-based approach. In: The 2005 ESK spring conference (2005)
12. Tanish, M.A.: Job process charts and man-computer interaction within naval command systems. Ergonomics 28, 555–565 (1985)
13. Dybkjær, L., Bernsen, N.O., Dybkjær, H.: Scenario design for spoken language dialogue systems development. In: the ESCA workshop on spoken dialogue systems, pp. 93–96 (1995)
14. Park, W., Han, S.H., Yang, H., Park, Y.S., Cho, Y.: A methodology of analyzing user input scenarios for a conversational interface. In: The 2005 ESK spring conference (2005)

Usability of Adaptable and Adaptive Menus

Jungchul Park, Sung H. Han, Yong S. Park, and Youngseok Cho

User Interface Laboratory, Department of Industrial and Management Engineering,
Pohang University of Science and Technology (Postech),
San 31, Hyoja, Namgu, Pohang, Kyungbuk, South Korea
{mozart,shan,drastle,kilys}@postech.edu

Abstract. This study investigates the usability of different adaptable and adaptive menu interfaces in a desktop environment. A controlled experiment was conducted to compare two different adaptive menus and one adaptable menu with a traditional menu. The two adaptive menus include an adaptive split menu that moves frequently used menu items to the top, and an adaptive highlight menu that automatically boldfaces frequently used menu items. Target selection times and the number of errors were recorded while the participants were performing menu selection tasks. Subjective satisfaction including perceived recognizability, perceived efficiency, and overall preference were also measured. The results showed that the adaptable menu outperformed the other menus in terms of both the performance and the satisfaction. The adaptive split menu was not as efficient as its theoretical prototype, especially when the selection frequency changed. The adaptive highlight menu, newly proposed in this study, was not significantly better than the traditional menu in terms of the selection time. However, it was preferred by the users since it helped them select frequently used items and was much less sensitive to the variations of selection frequency.

1 Introduction

Menu is one of the most important interface elements in a desktop environment. It becomes longer and longer as the number of functions in the software application increases. This stresses the importance of menu organization. Also, there is an increasing need for customization of menus, since computer systems today are designed to be used by millions of users with various purposes. There are two approaches to the customization/personalization of a menu. One is adaptable menu that can be modified by its users, and the other is adaptive menu that automatically adapts itself to the environment.

Some researchers found that making frequently selected menu items easy to select could reduce the selection time and increase the user satisfaction. Different adaptive menus have been proposed. Sears and Shneiderman [5] proposed 'split menu' in which three or four most frequently selected items are shown at the top of the menu. They compared it with an alphabetic menu and a frequency-ordered menu, and found that the split menu was better than the others in terms of performance and satisfaction when the frequently selected items were located in the middle or bottom of the menu.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 405–411, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Lee and Yoon [4] investigated the effects of various selection supports for menu selection tasks. They proposed a temporal selection support menu, which shows frequently selected items shortly before presenting the entire list of items. They compared it with the traditional, split and folded menus. The result showed that the split menu only provided spatial support by reducing the distance to high-priority items, whereas the folded menu provided both temporal and spatial supports. They showed that the split menu reduced the selection time in the early stage of use, and also was preferred by the users. However, the performance of the split and folded menus degraded as frequently selected items varied. The temporal menu was relatively insensitive to changes in selection frequency and location of the high frequency items, but there was no significant difference between the temporal and traditional menu in both the selection time and the user preference.

Findlater and McGrenere [3] compared three different types of split menus: (1) Static menu, in which four most frequently selected items always stay at the top of the menu; (2) Adaptive menu, in which two most frequently selected items and two recently selected items are presented at the top of the menu; (3) Adaptable menu, in which items could be moved to the top or bottom section of the menu by the user. The results showed that the static menu was faster than the adaptive menu, and the selection time of the adaptable menu was similar to the static menu in general, except when the adaptable was the first condition of the experiment, where it was as slow as the adaptive menu. The users preferred the adaptable menu to the static menu, but not to the adaptive menu.

The studies described above dealt with the quantitative comparison between various menu adaptation techniques. However, most of them are either incomplete or not systematic enough. Some of the studies are based on an unrealistic assumption about the system's knowledge of selection frequency. Some others that asserted the superiority of the split menu overlooked the adaptive menus' performance degradation under varying selection frequency. In addition, little effort has been made to examine the usability of adaptable menus, although it is generally accepted that many users want to adapt their system by themselves rather than fit themselves to the system.

This study examines the effectiveness of adaptable and adaptive approaches to menus in desktop applications. An adaptable menu and two different adaptive menus were implemented and evaluated. The two adaptive menus consist of an adaptive split menu that moves frequently used menu items to the top, and an adaptive highlight menu that automatically boldfaces frequently used menu items.

2 Methods

A controlled experiment was conducted to compare the usability of the menus. A total of 32 graduate or undergraduate students (25 males and 7 females) participated in the experiment. Their ages ranged from 17 to 25 (mean = 21.1) years. The participants were randomly assigned to one of two groups (Group A or Group B), each of which comprised 16 participants. The participants in a different group conducted experimental conditions in a different order of selection frequency distributions.

An application program was developed using Microsoft Visual Basic 6.0. Four menu types implemented in the program were traditional, adaptable, adaptive split, and adaptive highlight menus. The traditional menu is a typical static menu. In the adaptable menu, the user could move the menu items by dragging-and-dropping. The adaptive split menu is divided by a horizontal line into two sections. The top section keeps three most frequently selected items while the bottom section contains the others. It counts how many times each item has been used in recent 50 selections and updates the list after a selection is made. As in Sears and Shneiderman [5], both sections are organized in a traditional order that the entire list would have been presented in. In the adaptive highlight menu, the most frequently selected items are boldfaced instead of being moved. This menu selects the high frequency items in the same way as the adaptive split menu does. Fig. 1 presents the adaptable, adaptive split, and adaptive highlight menus.

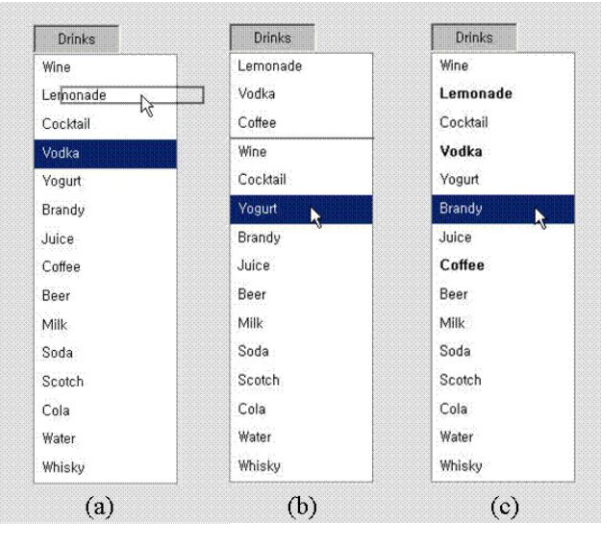


Fig. 1. Menu types: (a) adaptable menu (changing positions); (b) adaptive split menu; (c) adaptive highlight menu

Sixty different nouns from four label categories (fifteen nouns for each category) were used as labels of the menu items. The categories were body parts, countries, drinks, and sports. Nouns shorter than 4 characters or longer than 10 characters were excluded, and no more than four nouns in a category had the same initial letter. The category name was shown in the title bar at the top of the menu. Each menu contained 15 different items under a category.

The presentation order was counterbalanced using two Latin squares (one for the menu type and another for the menu label), which resulted in sixteen treatment conditions.

Two selection frequency distributions adapted from the literature with slight modification [4] [5] were employed to investigate the performance of these menus

under varying selection frequency. High frequency items are located in the top half in Distribution 1, while they are found in the bottom half in Distribution 2.

In the main experiment, the participants performed the four conditions (menu types) in the pre-determined order given by the experimenter. A condition was comprised of four task blocks, each of which contains 50 selections. The traditional menu was used in every first block of the conditions (Block 1) so that the participants could be acquainted with the label and occurring frequency of the items. Block 1 was excluded from the analysis. In the second, third, and fourth blocks (Block 2, 3, and 4), the participants were provided with one of the four menu types depending on the condition. The selection frequency started to vary at the beginning of Block 4. The participants in Group A, who had experienced Distribution 1 in the first three blocks, were presented with Distribution 2 in Block 4. Meanwhile, the participants in Group B, who had gone through Distribution 2 first, experienced Distribution 1 in the last block.

When a target item was presented on the screen, the participants were asked to open the menu by clicking the menu title, and then select the same item from the pull-down menu as quickly and accurately as possible. They were instructed to repeat the selections until they correctly selected the target item. When a participant selected the correct item, the menu was disabled for 1 second before the presentation of the next item. Target selection times and the number of errors (incorrect selections) were measured while the participants were performing the menu selection tasks. A short break (about two minutes) was given between the blocks. In the adaptable menu condition, the participants were given an opportunity to change the position of the menu items twice, once after Block 1 and once after Block 2. They could change the positions of the items if they wanted to do so. In a debriefing session, the participants were asked to rank the menu types on the basis of perceived recognizability of the items, perceived efficiency of selection, and overall preference.

3 Results

Means of the target selection times are presented in Table 1 and Table 2. Learning occurred in Block 1, and the selection time increased in most cases when the frequency changed in Block 4. For each participant group, the analysis of variance (ANOVA) was used to analyze the selection time for Block 2 and 3 combined (ideal performance), and Block 4 (performance under frequency variation) separately. Data from Block 2 and 3 were merged because the same condition was used in these blocks. The effect of menu types was found to be significant in all the four ANOVAs ($\alpha = 0.05$).

Differences between the menu types were analyzed by using SNK (Student-Newman-Keuls) test at $\alpha = 0.05$. The results showed that Group A participants were significantly faster with the adaptable and adaptive split menus than with the others in Block 2 and 3. However, in Block 4, the adaptable menu was the fastest, while the adaptive split menu was slower than any other menus. For Group B, the adaptable was significantly faster than the traditional and adaptive highlight in Block 2 and 3. There were no differences among the conditions except the adaptable. In Block 4, the

adaptive split was significantly slower than the others. But, there were no differences among the other conditions.

The number of errors per target selection ranged from 0.013 to 0.027 (0.019 on the average). Because the error occurred very rarely, a Chi-square test was used to analyze the number of errors. The result showed that there was no significant difference in the number of errors among the menu types (χ^2=2.509, p = 0.474).

Table 1. Means and standard deviations of selection time for Group A (s)

Block	Traditional		Adaptable		Adaptive split		Adaptive highlight	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
1	1.309	0.189	1.315	0.180	1.275	0.252	1.243	0.183
2	1.005	0.149	0.914	0.113	0.943	0.179	1.011	0.191
3	0.945	0.144	0.861	0.134	0.850	0.198	0.932	0.182
4	1.008	0.141	0.897	0.147	1.148	0.252	1.051	0.199

Table 2. Means and standard deviations of selection time for Group B (s)

Block	Traditional		Adaptable		Adaptive split		Adaptive highlight	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
1	1.390	0.289	1.393	0.237	1.390	0.282	1.367	0.329
2	1.082	0.227	0.964	0.146	1.066	0.214	1.079	0.255
3	0.983	0.178	0.902	0.181	0.908	0.186	1.002	0.230
4	0.925	0.152	0.925	0.208	1.162	0.198	0.968	0.226

The Friedman test revealed that, in terms of all the three subjective criteria, there were significant differences in rankings among the menu types at $\alpha = 0.05$. The differences were further analyzed using Dunn's post-hoc test at $\alpha = 0.05$. The participants considered the adaptable and adaptive highlight menus more recognizable than the others. As for the perceived efficiency, the adaptable was better than the adaptive highlight or traditional, whereas the adaptive split was only better than the traditional. The traditional was found to be perceived as less efficient than any other conditions. In terms of the overall preference, the traditional was the least preferred and there were no significant differences among the others.

4 Discussion

The traditional menu could be easily learned by the user, since it never changes the position or style of an item. If a user memorizes an item's position, he/she can easily find it. But, it usually takes time before he/she memorizes the position of all the items, and until then, he/she has no other choices but searching for the target item by scanning through the list. Even after he/she already knows the position of frequently used items, the menu does not provide any support to make the selection easier. This lack of support for finding and selecting frequent items explains why the traditional

menu was slower than the adaptable or adaptive split menus in Block 2 and 3. For the same reason, it was rated as the worst among the four by the participants. One half of the participants (50%) indicated that frequently used items were relatively difficult to select with this condition, and nine of them (28%) noted it was annoying when frequently selected items were located near the bottom of the list.

The adaptable menu was always one of the fastest. The participants could easily find target items by reordering the items or putting frequently used items at the top of the list. This leads to high rating scores in the perceived recognizability and efficiency by the participants. However, one shortcoming of this menu is that additional efforts are required for a user to reorganize the frequent menu items. Five participants (16%) indicated the inconvenience of performing adaptation on their own, and one participant (3%) did not adapt the menu at all because he felt it was unnecessary.

The adaptive split menu was significantly faster than the adaptive highlight or traditional menus when frequently used items were located in the upper half of the list and the selection frequency remained stable (Block 2 and 3 of Group A). In other conditions, it was not faster than the traditional one. A major drawback of this menu is that it is very sensitive to the variation of selection frequency. When frequently used items changed in Block 4, the selection time noticeably increased and it became slower than any other menu types regardless of the location of frequent items. It complies with the disadvantage of spatial inconsistency posed by Card [2] and Somberg [6]. According to them, positional consistency of items can increase the efficiency of search.

The adaptive split menu, which dynamically changes the items in the top based on the recent selection history, was found not to be as efficient as its theoretical prototype that has been highly rated in the literature. Although most participants (91%) responded it enabled efficient selection of items in the top, three-fourths of the participants (75%) complained that it was confusing when the selection frequency changed, and seven participants (22%) indicated the inconvenience of selecting infrequent items. For this reason, the adaptive split menu was not preferred to the traditional menu in terms of perceived recognizability, though it was rated high in terms of the perceived efficiency and overall preference. These imply that the utility of the split menu might be limited in practical applications where selection frequency is unstable and difficult to predict.

The adaptive highlight menu had no advantage in terms of the selection time compared with the traditional menu. It means that, in terms of the selection time, no benefit is obtained by highlighting frequently selected items, or the benefit, if any, is cancelled out by a side-effect. As four participants (13%) indicated, highlighted items sometimes interfered with the participants' search for a non-highlighted item. Nevertheless, the adaptive highlight was rated higher than the traditional menu in every aspect, and even higher than the adaptive split in perceived recognizability. Twenty five participants (78%) reported they could easily find an item when it was highlighted. Eleven participants (34%) indicated they were satisfied with this menu because the position of the menu items did not change.

The participants were able to efficiently organize the menu employing various criteria such as selection frequency and alphabetic order. A few participants used their likes and dislikes for the menu items or the word length. When they adapted the menu, they could easily find and select a target item. This advantage was maintained

even when the selection frequency changed. They seemed to recall the position of each menu item better, because they made the change by themselves. However, there remain some issues to resolve for practical applications. First, it is questionable whether the users are able to easily determine when the interface needs to be adapted to produce a benefit. Second, the users may not know what should be adapted to make menu selection more efficient. In this experiment, the participants adapted the menu after performing at least 50 selections for a short period of time. It may be difficult for the users to know the accurate frequency of each menu item or proper criteria for ordering the items, if they are asked to adapt the menu in the middle of long-term use. As Bunt et al. [1] suggested, providing adaptive support might help the users to determine what to adapt.

5 Conclusion

This study systematically compared the effectiveness of different adaptable and adaptive menus. The results have led us to conclude that the adaptable menus are very efficient and preferred by the users compared to the others. However, additional efforts are required for the users to adapt the menu, and it has been a major barrier that inhibits the adaptation. On the other hand, the adaptive split menu was not as efficient as its theoretical prototype, and the performance seriously deteriorated when selection frequency changed. These drawbacks seem to limit the utility of the split menu in practical applications. A new type of adaptive menus, i.e., the adaptive highlight was proposed and examined. This menu was preferred to the traditional one. Spatial consistency of the menu item in this menu type appealed to the users, though it could not reduce the selection time. Implications on the menu design were identified. The system needs to provide an efficient method for the users to easily reorganize the menu items using such criteria as frequency and recency of the items and alphabetic order. A hybrid combination of adaptive and adaptable menus would be interesting future research. The adaptive highlight might be applied to the adaptable menus to provide accurate selection frequency information based on history data and motivate the users to reorganize menu items at the same time.

References

1. Bunt, A., Conati, C., McGrenere, J.: In What Role Can Adaptive Support Play in an Adaptable System? International Conference on Intelligent User Interfaces, Madeira, Funchal, Portugal. ACM Press, New York (2004)
2. Card, S.K.: User perceptual mechanisms in the search of computer command menus. In: Proc. SIGCHI ’82, Gaithersburg, MD, USA, pp. 190–196 (1982)
3. Findlater, L., McGrenere, J.: A comparison of static, adaptive, and adaptable menus. In: Proceedings of the 2004 conference on Human factors in computing systems, Vienna, Austria (2004)
4. Lee, D.-S., Yun, W.C.: Quantitative results assessing design issues of selection-supportive menus. International Journal of Industrial Ergonomics 33, 41–52 (2004)
5. Sears, A., Shneiderman, B.: Split menus: effectively using selection frequency to organize menus. ACM Transactions on Computer-Human Interaction 1(1), 27–51 (1994)
6. Somberg, B.L.: A comparison of rule-based and positionally constant arrangements of computer menu items. In: Proc. SIGCHI/GI ’87, Toronto, Ont., Canada, pp. 255–260 (1987)

Towards Detecting Cognitive Load and Emotions in Usability Studies Using the RealEYES Framework

Randolf Schultz, Christian Peter, Michael Blech, Jörg Voskamp, and Bodo Urban

Fraunhofer Institut für Graphische Datenverarbeitung Rostock, Joachim-Jungius-Straße 11, 18059 Rostock, Germany
{randolf.schultz,christian.peter,michael.blech,joerg.voskamp, bodo.urban}@igd-r.fraunhofer.de

Abstract. In this paper, we will discuss some extensions to the RealEYES framework that can help to automatically detect interesting sections in usability studies using additional sensor input and knowledge discovery techniques.

Keywords: usability, emotions, cognitive load, human performance monitoring.

1 Introduction

Usability test systems usually collect a huge amount of various data. Screen recording, gaze tracking, mouse and keyboard input are just basic components of data streams nowadays usability test systems generate. Further data sources like audio data and face monitors add to the ever growing wealth of data to be processed and even newer technologies like emotion detection and human performance monitoring emerge and promise to add to the quality and value of usability studies.

Processing and analysing those data is very expensive in time and human resources, even if good tools for visualisation and analysis of the data are available. This is because common tools cannot find interesting sections in the test, where the subject experienced high mental load or prominent emotions, automatically. Instead, the usability expert has to browse manually to supposedly critical positions in the test data for analysis. Consequently, critical spots not envisioned by the expert may be missed.

The goal of the work presented here is to apply automatic analysis algorithms for identifying ongoing emotions and high cognitive load in the user to speed up the analysis process and spot critical situations in the data stream more easily. Commercially available tools do not offer equivalent functionality and are not extensible in a way we would need it. Consequently, we will discuss actual extensions to the RealEYES framework that help to automatically detect critical situations in usability studies by use of novel sensors and knowledge discovery techniques.

This paper is organized in the following way: first, the original RealEYES framework will be introduced briefly, followed by a description of the extensions for detecting emotions and cognitive load. We close with a discussion of exemplary study results and give directions for further work.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 412 – 421, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 The RealEYES Framework

The RealEYES framework combines a number of tools to support the entire process of a usability study, from preparation, execution, and analysis to communicating test setup data, measurement data and test results in an efficient manner.

The previous version of the framework already supported a multitude of data types in its data backbone: meta data on the test, screen-shot videos and video capture of the user, audio data, gaze and mouse pointer positions, and application and test specific events. The data is collected and synchronized in a single datastream per subject. After recording, the test data of the subjects may be visualized or analyzed in manifold ways including playback and statistical analyses.

2.1 The Structure of the Framework

The RealEYES framework consists of the following components: Recorder, Analyzer, Statistics, and Questioning (see Fig. 1).

They all work on a common information backbone comprised of AVI and XML files. The AVI files contain video and audio streams as well as gaze, mouse, keyboard, and event data, and the XML files contain meta data and intermediate test results. Thus, no lengthy conversion of data is needed when working with the tools in the RealEYES framework. All tools are equipped with sophisticated export functions for an easy integration of the test results into a usability report.

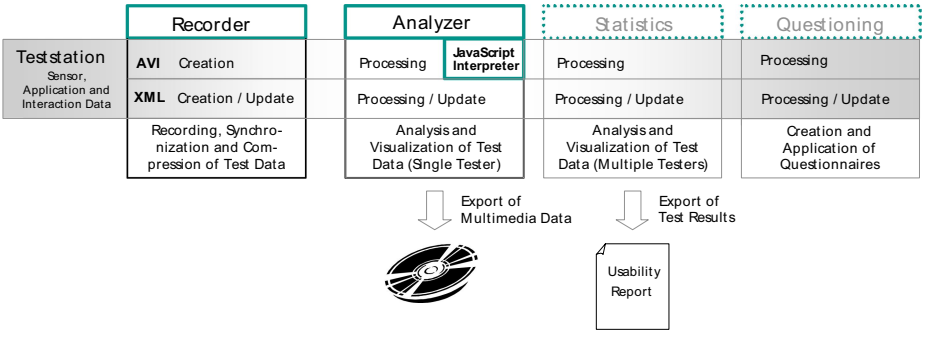


Fig. 1. Structure of the RealEYES framework

2.2 The Tools of the Framework

The most important tools in the RealEYES framework are Recorder and Analyzer. The *Recorder* manages the recording of all data. It requests the meta data, captures, synchronizes and compresses the test data and writes it to a single AVI file per session. The *Analyzer* is the main analyzing application that replays all video streams and visualizes the other data. Many standard and advanced visualizations of the screen-shot video together with the gaze and mouse data are available, such as a temperature grid; compare Fig. 4. The adjustable visualizations offer detailed insights into the actual user interactions with the tested product. New visualizations may be implemented by an advanced user of the system and plugged into the Analyzer.

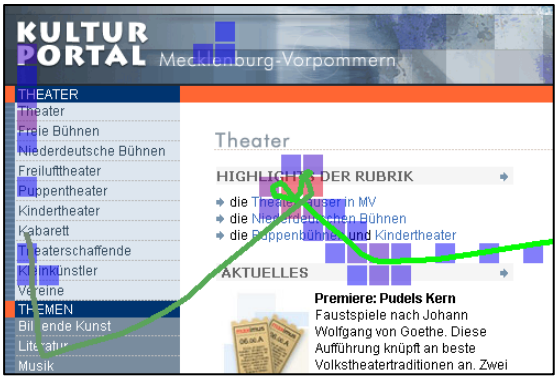


Fig. 2. Temperature Grid Visualization

The Analyzer is a tool for offline analysis of the collected test data and supports navigation using standard video player controls as well as event marks (see lower part of analyser window shown in Fig. 2). Event marks may be inserted automatically by the Recorder, manually by the usability expert during the test and the analysis phase, or semi-automatically using scripts written in JavaScript that run in the Analyzer's script engine during offline analysis. Regions of interest may be defined and simple statistical calculations can be performed on them. The regions may not only be defined geometrically but can also be bound in time or to certain tasks. Other tools of the framework may resort to these regions.

To improve the ability to test web-applications that often do not fit completely on the screen, both Recorder and Analyzer support and obey a scrolling region in all their features (see Fig. 4).

The two other tools in the RealEYES framework are Statistics and Questioning. The *Statistics* tool allows for complex statistical analyses to be accomplished on the acquired data (see Fig. 7). Furthermore, the Statistics tool is able to analyze and visualize data from all sessions of a study, i.e. to perform analyses over data of different users of a study. To illustrate, the Statistics tool can answer questions like "Did the majority of the subjects see the navigation buttons" or "What's the average time users looked at the advertisement". The *Questioning* tool allows the usability expert to create and utilize online questionnaires. The data gathered from the questionnaires is written to the XML file and can be processed by the Statistics tool.

3 Extensions of the RealEYES Framework

To further improve the frameworks capabilities, new technologies have been tested. Since user experience is closely coupled with emotions and cognitive load, sensors measuring related physiological parameters have been tried and finally incorporated into the framework.

For reliably detecting emotional states, various modalities can be used ([13], [7]). While face data of the user are already available in the data as well as speech, analyzing them is still a challenge (cf. [4]), particularly when it comes to facial feature analysis of subject thinking aloud. Progress has been made in recent years

particularly in speech analysis for emotional and cognitive signs (see also [4], [15]) and also in facial feature extraction ([5]; [1]).

The modality most researched and best understood today in terms of emotion is physiology, so we decided to add physiology sensors as well. After having examined various commercial sensor systems and decided against them for their inappropriate sensing elements and wires, we opted for the EREC sensor system developed by Fraunhofer IGD Rostock (see [11] and [12]). The RealEYES framework has been extended to support the open EREC protocol and Analyzer and Statistics tools have been extended to visualize and analyze the additional data.

3.1 Sensor Hardware

The additional sensors to gather physiological data should be minimally intrusive and easy to use. The EREC system (cf. [11]) developed at Fraunhofer IGD Rostock is a first step towards this. It features sensors in a fingerless glove, not hindering human computer interaction, built-in reliability checks, and wireless data transfer. The system has been tested in several studies and has been continuously improved (see [12]). Currently, heart rate, galvanic skin response and skin temperature are measured and made available immediately to the recording tool of the RealEYES framework. The sensor system is equipped with special error detection and reliability checks, which makes inclusion of physiological data into the processing application fairly easy. Particularly, EREC's convenient way of providing the data in engineering units made it very easy to incorporate the data without the need of implementing proprietary conversion algorithms. Also, with EREC implementing the SEVA standard for self-validating sensors ([6]; [3]), sensible data (from the processing point of view) are always provided, equipped with a standardized reliability flag for fast and easy appraisal of the data (see Fig. 3). For more details on the EREC system, refer to [11] and [12].

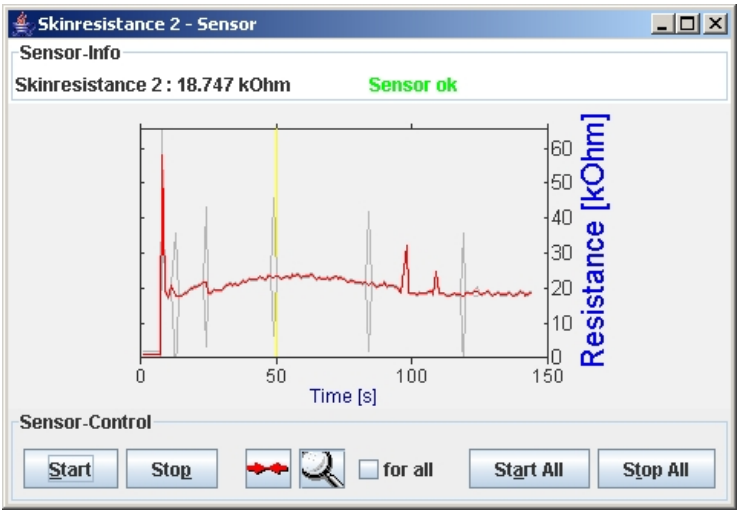


Fig. 3. EREC data are provided with sensor state and reliability information. The double spike zones indicate spots of uncertainty. Note that valid data are available for processing anyway. The line at 50ms represents an event mark.

3.2 Emotion Detection on Physiology Data

Sensing emotion-related information delivers a huge amount of data, with an enormous quantity of parameters being extracted from it, cf. [2]. At the current state of research, it seems to be necessary to collect physiological data at a relatively high sample rate. From our experience, 20 measurements per second are a good choice. This leads to 1200 samples per device and minute, accumulating to 36000 measured values per device for a half-hour experiment. To process the data, we apply knowledge discovery techniques because they allow to deal with big data sets and to examine them without previous specification of hypotheses or parameters to use. The extracted information can then be used to define attributes, statistical methods, learn algorithms, and classification concepts for integration into emotion detection classifiers, cf. [14].

Corpus. Based on a corpus previously build at Fraunhofer IGD Rostock (see [8]), filter operations and classifiers for emotional and cognitive states have been developed. They can be used to derive emotional states and cognitive overload from physiological data (heart rate, skin resistance and temperature). Those classifiers are to be integrated into the analyzing tools of the RealEYES framework.

In addition to the recording of physiological data, speech records are suitable to recognize current emotional states, as well. The research in emotion recognition from speech shows that the extraction of acoustic and prosodic features in combination with machine learning techniques leads to robust classifiers with success ranges from 60-80%, see [9].

To get more experience, we started an own study with an already existing speech corpus which is taken from the speech database of the TU-Berlin (cf. [10]). From this corpus we took 500 samples from male and female actors who have spoken several times a set of simple sentences, each time with an emotional intention from seven different categories.

In a first step, filters were applied that extract basic acoustic features such as duration, pitch, intensity and frequencies. In the following step, statistical calculations were applied on these basic features to characterize the current sample. Finally, the combination of the acoustic features with common statistical variables (min, max, median,…) and further typical speech processing values (longest voice, relation speech to non speech,…) leads to about 70 valuable features. The summary of these statistic features over all 500 samples is the basis for the final machine learning process. After performing tests with several different classifiers, we have reached a prediction performance of 76% with a support vector machine (compare Table 1).

The gained know-how for emotion detection from speech will be used in our next studies to improve the evaluation and interpretation of the study results. To be sure, to get valuable speech from a study, a test person is invited to give comments while performing a test. In addition, our speech corpus will be extended with the speech samples from new studies, so that the prediction performance and robustness of our speech classifiers can increase with every new study.

Table 1. Prediction performance on physiological and speech corpus

Physio.	*Best Classification Result*	*Random Classifier*
Max	49% (Euphoric)	20%
Min	26% (Helpless)	20%
Typical	38%	20%
Speech		
Typical	76%	14%

Store Study Results. To deal with the huge amount of data resulting from different studies, we have implemented a database to store emotion-related multimodal data, from physiological or speech recording modalities. The database scheme is designed to hold data from a (theoretically) unlimited number of sessions, of different studies, using any combination of input modalities. With this database, studies can be compared to other studies and data can be classified more precisely.

3.3 Integration into RealEYES Tools

To integrate the aforementioned sensing and emotion detection technologies into the RealEYES tools we decided to exploit the various extension methods already provided by the framework. Those are visualizations in the Analyzer and Statistics tool and events in the Analyzer.

Display of Raw Physiological Data. We started out with a graph display of the raw physiological data in the Analyzer (compare Figure 4). Using the graph display of the raw data, rapid changes in the physiological data (e.g. skin resistance) hinting stress may be discovered easily by the usability expert. The screen space available for this visualization, unfortunately, is not high enough to give a good overview of the data acquired during a complete test session. However, a rapid change is also easy enough to compute with JavaScript, leading to a script for the semi-automatic insertion of event marks. The event marks calculated by the script cover a complete test session and make it easy to navigate to all supposedly critical sections in the test. Nevertheless, we also plan to integrate a visualization of the raw data in graph form also in the Statistics tool. Here we could calculate and display values for a complete session and even for all test subjects from all sessions of the test at once. It would also be possible to create event marks in our data streams for exploitation in the Analyzer. This could be done even for situations that expose no conspicousness in the data of a single test subject.

Display of Classification Results. A visualization of classification results was already created in the context of our EmoTetris study (see [8]). We displayed a comic face to visualize a predominant detected base emotion and used a star plot diagram to visualize the results of all classifiers for base emotions (see Figure 5).

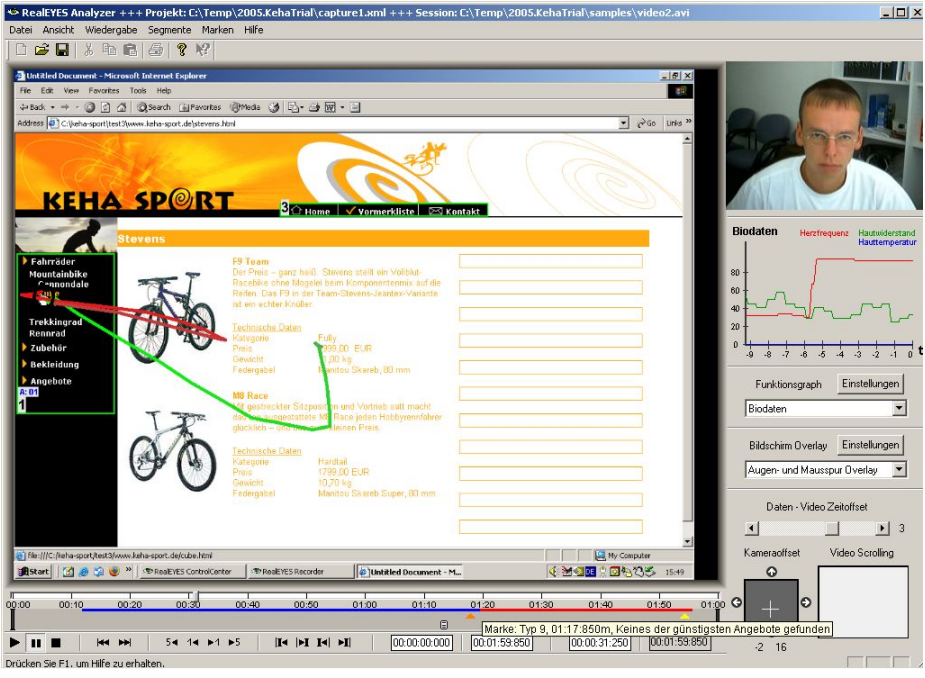


Fig. 4. RealEYES Analyzer Tool with Visualization of raw physiological data (below the image of the test subject) and Event Marks (below the time line)

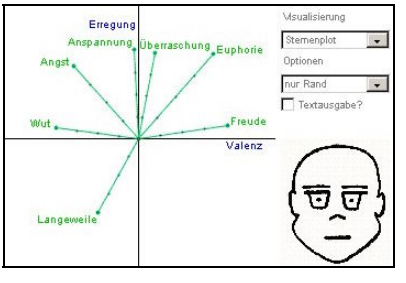


Fig. 5. Visualization of Emotion Classification Results in EmoTetris; Star Plot of Base Emotions in Russel Diagram (*left*), Comic Face (*lower right*)

Since this visualization only shows a current state and no process, it is not very well usable in the RealEYES context and we plan to integrate a graph display similar to the graph display for raw physiological data into the Analyzer and the Statistics tools instead. Similar to the script that detects rapid changes in physiological data, event marks will be inserted when a classifier detects e.g. a high stress level.

Measures must be taken to not flood the data stream with events, especially when classifiers for different modalities are active. Further problems arise when the classifiers deliver contradicting results. Those problems may be solved by creating a second classifying layer that works on the output of the classifiers for different modalities.



Fig. 6. Conventional Heat-Map (*left*) and Read-Map (*right*) of two test subjects in comparison

In our EmoTetris study we discovered that apart from the basic emotions, states like loss of control also play an important role in the human computer interaction. Other states that are relevant in the context of usability are searching and reading. Consequently, we developed classifiers for those states and integrated them in the Statistics tool. The classifiers work solely on gaze and mouse data as input. The output of the classifier may easily be mapped to the screen-content of the recorded application guided again by the gaze data. The result is an enhanced version of a Heat-Map, visualizing parts on a screen image that have been read (rather than just looked at) by the subject (see also Figures 6 and 7), hence the name Read-Map. While the left image of Figure 6 only shows that the test subjects looked at the text sections on the underlying web page, the right image makes it clear that the texts actually have been red. For usability studies, Read-Maps give valuable hints for actual reading activity (an interesting section of a usability test) on e.g. E-Learning content, or advertisements.

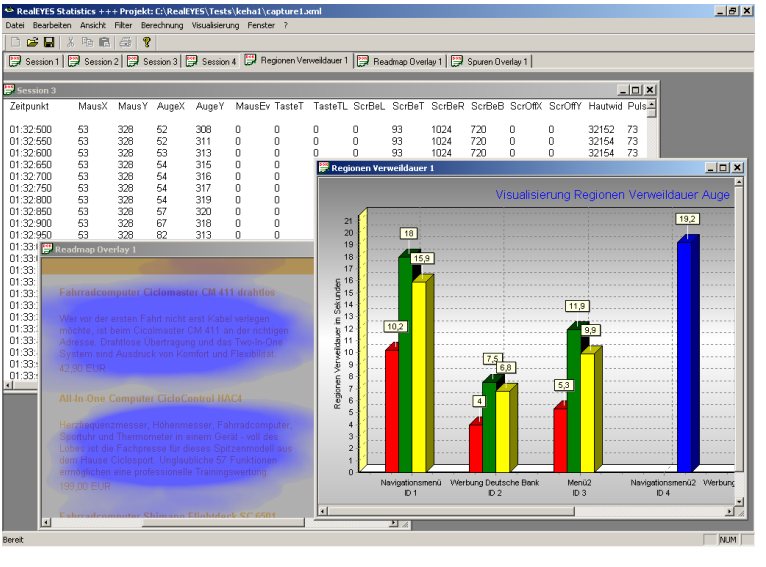


Fig. 7. User Interface of RealEYES Statistics with ReadMaps on the text display

4 Conclusions

In this paper we discussed some extensions to the RealEYES framework that allow to integrate and process sensor data with machine learning algorithms to automatically find critical incidents in usability studies. We presented some detection results, showing that our approach is not only applicable but opens new perspectives in usability studies. Further work will focus on improving the detection rates using more physiological parameters (breathing activity, oxygen level) and improved knowledge discovery algorithms. We would also like to integrate another in-house-developed tracking system for human motions to extend our scope beyond the desktop scenario.

References

1. Aleksic, P.S., Katsaggelos, A.K.: Automatic Facial Expression Recognition Using Facial Animation Parameters And Multi-Stream Hmms. In: IEEE Trans. on Sig. Proc. Supplement on Secure Media (2005)
2. Blech, M., Peter, C., Stahl, R., Voskamp, J., Urban, B.: Setting up a multimodal database for multi-study emotion research in HCI. In: Proceedings of the HCI International Conference, Las Vegas (2005)
3. BSI, 2004. British standards institute: BS 7986: Industrial process measurement and control – Data quality metrics (2004) Available from BSI Customer Services email: orders@bsi-global.com
4. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human computer interfaces. IEEE Signal Processing Magazine (January 2001)
5. Fasel, B., Luettin, J.: Automatic Facial Expression Analysis: A Survey. Pattern Recognition 36(1), 259–275 (2003)
6. Henry, M.P.: Self-Validating Sensors – Towards Standards and Products. Automatione e Strumentazione (2001)
7. Hudlicka, E.: Affect Sensing and Recognition: State-of-the-Art Overview. In: Proceedings of the 2005 HCI International Conference, Las Vegas. vol. 11 (2005) CD-ROM ISBN 0-8058-5807-5
8. Oertel, K., Schultz, R., Blech, M., Herbort, O., Voskamp, J., Urban, B.: EmoTetris for Recognition of Affective States. In: Proceedings of the 2005 HCI International Conference, Las Vegas (2005) CD-ROM. ISBN 0-8058-5807-5
9. Oudeyer, P.: The Production and Recognition of Emotions in Speech: Features and Algorithms, Sony CSL Paris (2003)
10. Paeschke, A.: Prosodische Analyse emotionaler Sprechweise, Logos Verlag, Berlin (2003)
11. Peter, C., Ebert, E., Beikirch, H.: A Wearable Multi-Sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In: Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction, Beijing, 2005, pp. 691–698. Springer Verlag Berlin, Heidelberg, New York (2005)
12. Peter, C., Oertel, K., Kaiser, R., Schultz, R., Göcke, R., Voskamp, J., Urban, B.: The EREC sensor system for affect detection - application studies and results. Special session on emotion in HCI at the HCI International Conference, Beijing (2007)

13. Picard, R.W.: Affective Computing for HCI. In: Proceedings of the 8th International Conference on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I, Lawrence Erlbaum Associates, Inc., Mahwah, NJ (1999)
14. Picard, R.W., Vyzas, E., Healey, J.: Toward Machine Emotional Intelligence - Analysis of Affective Physiological State. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23(10) (October 2001)
15. Tonti, M.: The influence of emotional and cognitive processes in the definition of speech rate. 37th annual meeting of the Society for Psychotherapy Research. Edinburgh (June 21–24, 2006)

Relationship Model in Cultural Usability Testing

Qingxin Shi and Torkil Clemmensen

Department of Informatics, Copenhagen Business School, Denmark
{qs.inf,tc.inf}@cbs.dk

Abstract. Culture plays an important role in the global market today. It not only affects products, but also impacts on usability evaluation methods. In this paper we first introduce culture theories and two kinds of relationships in thinking aloud usability testing and then review previous research. Based on the discussion, we extract the potential factors which may influence cross-culture usability testing and then propose a relationship model. Finally, we discuss how the two thinking aloud approaches may be used in cross-culture usability testing.

Keywords: Usability test, culture, thinking aloud theory, localization.

1 Introduction

With the advent of globalization and IT revolution, we can no longer overlook the aspect of culture in the design of user interfaces and products. In order to capture global markets, the products and software must be tested in target cultures to make sure that they are acceptable and suitable for people's cultural characteristics. But some previous studies have found that culture not only influences the products or interface design, but also the design methods used in building interfaces [1]. Culture affects the usability evaluation methods (UEMs) of focus groups, questionnaires, structured interviews, and the understanding of metaphors and interface design [2]. Yeo's study [3] shows culture also impacts on usability testing.

Thinking aloud usability testing, has been extensively applied in industry to evaluate a system's prototypes of different levels of fidelity [4]. The primary goal of a usability test is finding a list of usability problems from evaluators' observations and analysis of users' verbal and non-verbal behavior; thus, the relationship between the evaluator and user is very important for finding accurate usability problems. Tamler [5] suggested establishing a trusting and supportive relationship in order to make the users honestly disclose their thoughts and feelings.

During usability testing, representative users are required to complete pre-established tasks by using the system. This measurement is largely related to specific users and specific tasks. However, people differ across regional, linguistic and country boundaries; therefore, if the evaluator and user have different cultural backgrounds, they may be strongly influenced by their local cultural perspective, perception and cognition, so the interaction and communication between them may be different from those who are from the same culture. Since usability testing involves human-human interaction, the evaluator and user's cultural background must be

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 422–431, 2007.
© Springer-Verlag Berlin Heidelberg 2007

considered, or else there may be a misunderstanding between them. Therefore, how to build an effective relationship in the usability test has become a key issue in cross culture usability testing.

Although the thinking aloud usability test is generally thought to be an effective and successful technique [6], practitioners do not conform to the theoretical basis of the thinking aloud method in the industrial area which was described by Ericsson and Simon [7]. Therefore, Boren and Ramey [7] proposed speech communication theory as a theoretical basis for thinking aloud in usability testing, focusing on evaluator-test user communication. Later we will introduce two kinds of relationships in usability testing using the two thinking aloud theories.

Based on previous research, this paper analyzes the concept of cultural usability testing and brings forward a relationship model in cultural usability testing.

2 Two Kinds of Relationships in Usability Testing

Thinking aloud has been a widely used method to study people's thought processes and content. There are two different thinking aloud theories which can inform us about the evaluator and test user relationship in usability testing. One is the classical thinking aloud method which was put forward by Ericsson and Simon. This classical model emphasizes that during the usability test session, there should be little interaction and communication between the user and evaluator. The evaluator just tells the user to speak "as if alone in the room" [7, p.263]. The only interaction may be asking the user to keep thinking aloud. There is no tight relationship between the evaluator and test user.

The other thinking aloud theory in usability test was brought forward by Boren and Ramey [7] called speech communication theory. This theory focuses on the communication between the evaluator and test user. In the practice of usability testing, there is always a user and an evaluator. "Talk is not simply a form of action" performed by the user alone, "but a mode of interaction" between users and evaluators [7,p.267]. Relationship is much more important in this theory. The users cannot ignore the evaluators, even silent ones. They expect a response, agreement, sympathy, etc., from the evaluators. In the speech communication model, some key issues need to be clarified:

- The subject of the test is the interface, not the user.
- The test user is the expert, who is assumed to provide valuable information of the interface. The evaluator is the learner, whose main task is to get information from the user's speech and find usability problems.
- The evaluator should use undirected and undisturbed tokens to keep the users focused on the tasks and at the same time, verbalize their thoughts fluently.
- When encountering contingencies during the usability test, interaction between the evaluator and test user is required.
- In the practice area, it is okay to probe with questions to elicit more valuable information, which is not allowed in Ericsson and Simon's theory.

Ericsson and Simon's theory is primarily focused on cognitive processes, like problem solving. However, in the usability test, the main purpose is not only to get

the user's thoughts, but more importantly, to get the user's expectations, feelings, design ideas, etc., of the interface/software. So as long as the evaluator does not force his/her own opinion on the user, it is okay to build a proper interactive relationship between the evaluator and test user in the usability test. We did some field studies in Denmark, India and China. We observed how the usability practitioners do the usability test in the industrial area. Actually in all three countries the evaluators did not listen passively, but actively interacted with the users on necessary occasions to get more valuable information about the interfaces.

We can say that having a good communication and interaction is very necessary for usability testing. In order to get the fluent communication and suitable interaction, a warm, supportive and trusting relationship cannot be ignored.

3 Culture and Culture Theory

Culture has been defined in many different ways by different researchers. With regards to the usability test, we need a more specific definition of culture. Thus we introduce Honold and Nisbett's conception of culture. Honold [8] defines culture for the purposes of human computer interaction. One of her definitions is worthy of mentioning:"Culture does not determine the behavior of individuals but it does point to probable modes of perception, thought and action. Culture is therefore both a structure and a process" [8,p.329]. From Nisbett's research, we understand that people in different cultures perceive the world differently, which means that people's cognition and perception are different in different cultures. "Cultural practices and cognitive processes constitute one another. Cultural practices encourage and sustain certain kinds of cognitive processes, which then perpetuate the cultural practices" [9, p.3]. Usability testing is a cognitive activity [10] which, from the evaluator's eye, sees the user's behavior and comments. If they are from the same culture, it may be much easier for the evaluator to get the user's real meaning. If using a foreign evaluator, it will require extra effort to understand the user's real meaning. Hence, the effective communication and interaction is much more important in a cross cultural usability test.

Regarding culture theory, considerable usability research cites Hofstede [2, 3, 11-13] who has proposed five culture dimensions: power distance (PD), collectivism/individualism (IC), femininity/masculinity (MF), uncertainty avoidance (UA), long-term Confucian orientation. Marcus [11] has investigated how culture dimensions might affect user-interface designs. His research seeks to help user-interface designers cope with global product and service development. Although it is hard to design a universally usable interface, it is possible to provide guidelines for UEMs applied in different cultures.

This paper elaborates on Nisbett's culture theory [9, 14]. His theory focuses on the cognition and perception differences; for example, people from western countries and eastern countries will be different in causal attribution, categorization, and attention to the context vs. salient object [15]. This theory is more relevant to usability testing because thinking aloud usability evaluation methodology asks users to work on typical tasks and to verbalize their task performance and thought process [16]. The whole process involves users' cognition and perception characteristics. The results of the

usability test, i.e., usability problems, which are found by the evaluators, are also involved in the evaluators' cognition and perception of the whole test process. When cultural differences exist between the evaluator and test user, some usability problems might be masked, instead of being uncovered. If the cultural influence is ignored, the usability test UEM methodology may be inefficient to provide accurate information about the localized product.

From Nisbett's culture theory, there are two kinds of orientation [17]: task-focus orientation and socio-emotional relational orientation. Task-focus orientation means people's effort is directed towards task-related goals, and attention is focused on monitoring the extent to which these goals are being accomplished. Socio-emotional relational orientation means people's effort and attention are directed towards the interpersonal climate of the situation, and they strive to maintain social harmony. Users from different cultures may be affected by the foreign evaluator/interviewer in quite different degrees. Users, from socio-emotional relational orientation cultures, may be influenced more by the perception of a foreign evaluator. On the other hand, users from a typical task-focus culture may not be influenced by the foreign evaluator, since they focus only on their task and do not care much about the evaluator's status.

4 Previous Work on Relationship in Cultural Usability

Yeo [3] examined cultural factors that may affect the results of usability evaluation techniques. The aim of his study was to identify, examine and reduce the effect of cultural factors that influence usability testing. Initial results showed that an important possible cultural factor is power distance: a test user who was of higher rank than the experimenter gave more negative comments about the product than the one who was of lower rank than the experimenter. Nisbett's culture theory suggests that Malaysian culture is socio-emotional relational orientation culture. In the usability test the users hope to establish a harmonic relationship with the evaluators, so they do not want to give too many negative comments during the usability test, even if it is very hard for them to complete the task using the system. If the user thinks the evaluator has a higher rank, they may be more reluctant to provide negative comments [3] since they do not have a task-focus orientation; rather, they hope to build a good relationship with the higher ranking evaluator. So in Malaysian culture, in order to get honest results from usability testing, the experimenter should be of the same rank or of lower rank than the test subjects.

Yeo [13] explored the efficacy of the global-software development lifecycle (global-SDLC), which includes the design, implementation and usability evaluation phase. He found that adapting software from a source culture to a target culture, the design and implementation phase is efficacious, but the evaluation phase is not. He employed three usability Assessment Techniques (UATs): Thinking-aloud Technique (objective measure), System Usability Scale (subjective measure) and Interviews. The results of the usability evaluations were found to be inconsistent. He found that for the less experienced computer users, or for the users who were not familiar with the evaluators, the objective measure and subjective measure were not matched. Even though these users performed poorly on the task, they still provided positive comments of the software in the interview. According to Yeo, the cause of these

inconsistencies was the users' reluctance to provide critical negative comments. Malaysia is a collectivistic country where users want to 'preserve the face' of the designer. If Malaysian users were familiar with the evaluator, they would not be concerned about making negative comments [13] since they would understand the evaluator's role in the usability test, and know that their negative comments would not destroy the good relationship with the evaluator.

Vatrapu and Pérez-Quiñones [2] investigated the effects of culture on structured interviews in the usability test. They carried out controlled experiments using two independent groups of Indian participants by two interviewers. One interviewer was from Indian culture and the other from Anglo-American culture. The results showed that the culture of the interviewer had an effect on the number of usability problems found, on the number of suggestions made, and on the number of positive and negative comments made. They found that the participants who were from the same culture as that of the interviewer (Indian culture) brought more usability problems and made more suggestions than participants who were interviewed by the interviewer who was not of the same culture (Anglo-American). From their study, we can see that when using a foreign evaluator, users may not be willing to talk as freely and accurately as when using a local evaluator. Language may not be the key issue, since in this research both interviewers and users could speak English fluently. We will analyze the potential factors that may influence cross-cultural usability testing.

5 Main Factors in Cultural Usability Testing

From the above discussion related to thinking aloud theories, culture theories, and previous research, we have extracted the basic factors that may influence cross-cultural usability testing. We will briefly discuss these factors now.

5.1 Evaluator and User's Cultural Background

Cultural background needs to be considered since users from different cultures may not be influenced to the same degree when they are with a foreign evaluator. Sanchez-Burks's study [17] found that Northern European culture is a typical task-focus culture, which means that users in those countries may not be influenced when the evaluator is from another country since they pay more attention to the task, not the evaluator. While East Asian culture and Indian culture are socio-emotional relational orientation cultures, users in these countries may be influenced more when they are with a foreign evaluator. For example, the study done by Vatrapu and Pérez-Quiñones [2] shows that Indian users who were with a foreign evaluator did not like to talk as freely as those who were with a local evaluator. But this may not be the case for Danish users. In our future study, we will use a foreign evaluator in India, China and Denmark to see whether the effect degree is the same in different kinds of cultures.

5.2 The Application/Software/Interface Being Tested

The requirement of an evaluator's cultural background is also related to the application or product which is tested in the target culture. There are two approaches to designing products for international markets: globalization and localization [18].

"Globalization seeks to make products general enough to work everywhere and localization seeks to create custom versions for each locale" [18,p.158]. If testing a localized application which adapts specific cultural elements for a specific target culture [19], the results of the usability test may be more related to the evaluator and user's cultural background. Usability testing will not provide accurate information when a localized product is tested without considering cultural issues.

In Vatrapu and Pérez-Quiñones's study, the website which was tested was a culturally localized website, which means people in other cultures might not understand the background, purpose and other detailed issues of it. It is not easy for a foreign interviewer to find the culturally sensitive usability problems. On the other hand, the users also did not discuss too much with the foreign interviewer since they thought the foreign interviewer did not understand it. The users with the foreign interviewer just gave their opinions with little communication and interaction with the interviewer which, in turn, influenced the usability problems that the foreign interviewer would find. This implies that when testing a culturally neutral application, the influence of the difference in cultures between interviewer and user may not be as big as a culturally localized application. In our future study, if we want to see bigger cultural influences, maybe we should still use culturally localized application/software. Of course, we can also compare the difference of testing culturally localized applications and culturally neutral applications to see whether cultural issues have the same effect.

5.3 Evaluator Effect

The influence of culture on usability testing may also be derived from another factor called the Evaluator Effect: the total number of usability problems found will depend upon the knowledge and experience of the evaluator and the number of evaluators [6].

Hertzum and Jacobsen [10] examined three of the most widely used usability evaluation methods, cognitive walkthrough, heuristic evaluation, and thinking aloud, and found that all of them suffer from a substantial evaluator effect. No two evaluators evaluating the same interface and using the same usability evaluation method found the same set of problems. The evaluator effect exists "for both novice and experienced evaluators, for both cosmetic and severe problems, for both problem detection and severity assessment, and for evaluations of both simple and complex systems" [10,p.421].

The evaluator effect indicates that even in one culture, evaluators with different experience will find different usability problems. The effect may be much more significant when the evaluators are from two different cultures, since they do not even have the same cultural background. Even though they are both very qualified and professional, their cognitive process and knowledge cannot be the same, which may be a strong impact factor on cross cultural usability testing.

In a cross-cultural usability test, how can we minimize the evaluator effect which is derived from culture? It is very hard to change the foreign evaluator's cognitive process, but it may be much easier to increase his/her knowledge related to the culturally localized application. The foreign evaluator does not need to master all the target culture, because it is impossible. But he can get some important information just related to this application. Maybe he/she needs to know the background, using habits and some related culture features of the application in the target culture, which will be very helpful for them to understand and communicate with the users in the usability test.

6 Relationship Model in Culture Usability Testing

Based on the above discussion, a cultural usability testing model was brought forward (see Figure 1).

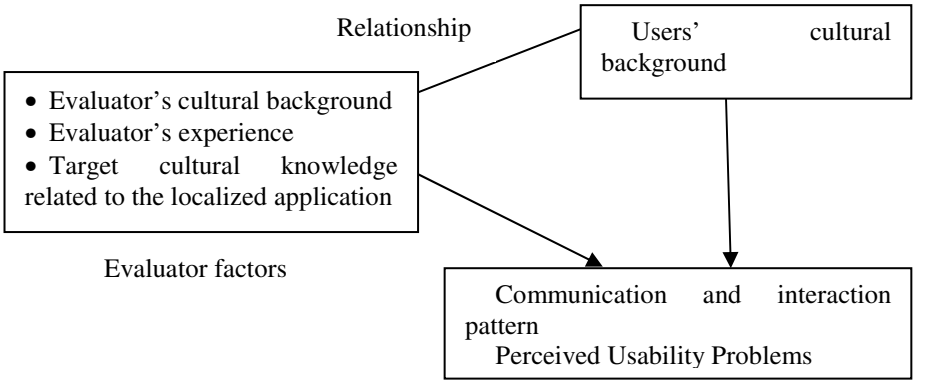


Fig. 1. Relationship Model in Cross-Cultural Usability Testing

This model considers the evaluator's cultural background (task-focus versus socio-emotional relational orientation), experience and knowledge related to the localized application. There are four basic relationships between the evaluator and test user:

1. Foreign experienced evaluator with little target culture knowledge about the localized application----local users.

In this model, all the participants speak English in the target country. The foreign evaluator just gets the instructions of the tasks and the procedures of the test, but does not have any training for the localized application. The foreign evaluator does not have much knowledge about the usage of the culturally localized application in the target culture. But the application also exists in the evaluator's culture and the evaluator is familiar with such application in his/her own culture; thus, the only knowledge the foreign evaluator would need to master is the related cultural issues. If the foreign evaluator does not know the application at all, for example, chopsticks are seldom used by Danish people, then it would not be necessary to ask a Danish evaluator to do the usability test with Chinese users in China.

In our pilot study [20], we used Microsoft Clipart as the application, since regardless of which culture the evaluators came from, they would know what Clipart is and how to use it. But the Clipart which was tested is a culturally localized one. We added a collection of culturally specific images and icons and a text document with preformatted invitation text called "cultural clipart" to My Collections in Microsoft Word's clipart organizer. The usability test is to see whether the "cultural clipart" is good enough for the user to make a traditional wedding invitation in the target culture.

Above all, in this model, the foreign evaluator should be usability professional. The application which is tested should be a common application but culturally localized. The evaluator does not have such knowledge about the localized application

related to the cultural issue. Our aim is to see what is the communication and interaction pattern in this situation, and how many and what kind of usability problems the foreign evaluator would find.

2. Foreign experienced evaluator with more target culture knowledge about the localized application ----local users.

This model is similar to the first relationship. The only difference is that the foreign evaluator will be trained with important related cultural information about the localized application.

3. Local experienced evaluator----local users.

In this model, all the participants speak English. If the communication and interaction is better than the first two models, then we can safely infer that the language does not have a great influence that the difference may be from the cultural background. Of course, in all the four models, participants are chosen who are good at English.

4. Local novice evaluator----local users.

This model is similar to the third, except that the local evaluator is not experienced.

In the four relationship models we can compare Model 1 and Model 2 to see the influence of knowledge on the results of the usability test (relationship built in the test; communication and interaction pattern; perceived usability problems). Comparing Model 2 and Model 3, what kind of knowledge does the foreign evaluator have to get in order to do the usability test as efficaciously as the local evaluator? Suppose in Model 2 the foreign evaluator mastered all the related information, and then compared it to Model 3, would they get the same result? If not, what are the other main factors that influence the cross-culture usability test? Comparing Model 3 and Model 4, what is the influence of the evaluator effect? Comparing Model 1 and Model 4, which factor is more important, knowledge related to the culture or the skill of doing a usability test. By making these four groups of comparisons, we hope to gain a clearer understanding of cross-cultural usability testing.

6.1 Two Thinking Aloud Theories in Cross-Culture Usability Testing

As introduced above, there are two different thinking aloud approaches in a usability test. The usability practitioners usually do not follow the rigid thinking aloud approach which was proposed by Ericsson and Simon. Tamler [5] claims that thinking aloud data which is generated by the users themselves is often inadequate. The evaluator needs to probe questions which are important for the interface but not noticed by the user, and he/she also needs to share his/her understanding of the user's speech and behavior and get feedback from the user in order to get the user's real idea and experience to the interface[5]. Therefore, the communication and interaction is very important for a usability test. A fluent and successful communication and interaction also relies on a supportive relationship. Actually, it is much harder for foreign evaluators to establish such supportive relationships with native users in the target country, even for experienced usability professionals.

When professional evaluators conduct a usability test with foreign users in the target culture, they may follow the traditional way (Ericsson and Simon's approach) to do the thinking aloud, which means that there might be less interaction and

communication. Since they may not be familiar with the culturally localized application, they may not be certain what the critical issue is that needs to be probed. In order not to disturb and influence the users and get more accurate information, the better way is to interrupt them less and avoid false leading.

When native professional evaluators conduct a usability test with native users, they may be following the communication theory proposed by Boren and Ramey [7]. Previous studies [2, 20] show that, compared to foreign evaluators, local evaluators had more interrogative reminders, affirmative reminders, and help out behaviors.

Krahmer and Ummelen conducted two variants of usability tests under controlled circumstances. One condition was based on Ericsson and Simon's protocol, and the other on Boren and Ramey's proposal [21]. From their research, they found that although the evaluators used different approaches, the process of thinking aloud while carrying out tasks is not affected by the type of approach that was used. The task performance does differ. More tasks were completed in the Boren & Ramey condition, and subjects were less lost. But the number of different navigation problems that were detected and users' evaluations of the website quality were similar.

From this study, we can see that no matter which thinking aloud theory was followed, experienced usability professionals will find similar usability problems. In a specific culture with a local evaluator, the users' evaluation of the application/software will not be influenced by the thinking aloud approach that is used in the usability test. If foreign and local evaluators find quite different usability problems, this may not be because they are following different protocols with local and foreign users, but may be due to their varied cultural background.

7 Conclusion

This paper has discussed the effects of culture on thinking aloud usability testing and the application of two thinking aloud theories in cross-cultural usability testing. As the usability test needs test organizers, evaluators and users who may be from different cultures, it is becoming increasingly important to avoid the effect brought about by cultural differences. In this paper we have discussed the cultural influence on usability testing from only a theoretical viewpoint. In future studies we intend to investigate from an empirical viewpoint what kind of relations and communications between evaluators and test users are the most effective for finding usability problems of a culturally localized application during the usability test. We will design experiments to validate the models proposed above.

Acknowledgments. This study was co-funded by the Danish Council for Independent Research (DCIR) through its support of the Cultural Usability project.

References

1. Vatrapu, R.: Culture and International Usability Testing: The effects of Culture in Structured Interviews. Master thesis, in Virginia Polytechnic Institute and State University (2001)
2. Vatrapu, R., Pérez-Quiñones, M.A.: Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. Journal of Usability Studies 1(4), 156–170 (2006)

3. Yeo, A.W.: Cultural Effects in Usability Assessment. In: CHI 98, Doctoral Consortium (1998)
4. Law, E.L.-C., Hvanneberg, E.T.: Analysis of Combinatorial User Effects in International Usability Tests. In: CHI 2004. Vienna, Austria (2004)
5. Tamler, H.: High-tech versus high-touch: The limits of automation in diagnostic usability testing. http://www.htamler.com/papers/techtouch/
6. Clemmensen, T., Goyal, S.: Cross cultural usability testing. Working paper, Copenhagen Business School, Department of Informatics, HCI research group, 2005-006, p. 20 (2005)
7. Boren, M.T., Ramey, J.: Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication 43(3), 261–278 (2000)
8. Honold, P.: Cultural and context: an empirical study for the development of a framework for the elicitation of cultural influence in product usage. International Journal of Human-Computer Interaction 12(3&4), 327–345 (2000)
9. Nisbett, R.E., Norenzayan, A.: Cultural and Cognition. In: Medin, D.L. (ed.) Stevens' Handbook fo Experimental Psychology, 3rd edn. (2002)
10. Hertzum, M., Jacobsen, N.E.: The evaluator effect: A chilling fact about usability evaluation methods. International Journal of Human-Computer Interaction 13(4), 421–443 (2001)
11. Marcus, A.: User Interface Design and Culture. In: Aykin, N. (ed.) Usabillity and Internationalization of Information Technology, pp. 51–78 (2005)
12. Vöhringer-Kuhnt, T.: The influence of culture on Usability. Master thesis. in Dept. of Educational Sciences and PsychologyFreie Universität Berlin (July 2004) 2002: Berlin, Germany, http://userpage.fu-berlin.de/~ kuhnt/thesis/results.pdf
13. Yeo, A.W.: Global-software Development Lifecycle: An Exploratory Study. In: CHI (2001)
14. Nisbett, R.E.: Cognition and Perception East and West. In: 28th International Congress of Psychology in Beijing (2004)
15. Nisbett, R.E., Masuda, T.: Cultural and point of view. PNAS, 2003 100(19), 11163–11170 (2003)
16. Ramey, J., et al.: Does Think Aloud Work? How Do We Know? In: CHI 2006 (April 22-27) (2006)
17. Sanchez-Burks, J., Nisbett, R.E., Ybarra, O.: Cultural Styles, Relational Schemas and Prejudice Against Outgroups. University of Michigan (2000)
18. Horton, W.: Graphics: The not quite universal language. In: Aykin, N. (ed.) Usabillity and Internationalization of Information Technology, pp. 157–188. Lawrence Erlbaum, Mahwah (2005)
19. Bourges-Waldegg, P., Scrivener, S.A.R.: Meaning, the central issue in cross-cultural HCI design. Interacting with Computers 9(3), 287–309 (1998)
20. Clemmensen, T.: Cultural models in psychological usabililty evaluation methods (UEM). In: Indo-Danish HCI Research Symposium (2006)
21. Krahmer, E., Ummelen, N.: Thinking About Thinking Aloud-A comparison of two verbal protocols for usability testing. IEEE Transactions on Professional Communication 47(2), 105–117 (2004)

An Empirical Evaluation of Graphical Usable Interface on Mobile Chat

Victoria Yee Siew Yen[1] and Daniel Su Kuen Seong[2]

[1] Financial Services, Accenture, Malaysia
victoria.yee@accenture.com
[2] The University of Nottingham, Malasysia
daniel.su@nottingham.edu.my

Abstract. Current text-based mobile group chatting systems hinder navigation ease through long chat archive in a limited screen display. Moreover, tracking messages sent by specific chatter is cumbersome and time consuming. Hence, graphical-based usable interface that aids navigation and message tracking through minimal key-pressed and enhances user expression via avatars employment is proposed. The research outcomes typified that there was significant linear relationship between user interface and usability on text-based and graphical-based usable interface on mobile chat. Moreover, the experimental evaluation results indicated that text-based usability could be improved by creating interface that encourages usages whereas the graphical-based usable mobile chat is augmented by crafting user friendly interface that enhances user satisfaction, encourages usages and promotes navigation ease. The empirical findings and results exemplified that the potential use of graphical-based usable mobile chat as substitution to the text-based that presently has poor reception and is under utilised in commercial arena.

1 Introduction

Communication is a salient part of life and is used to convey social presence, augment social bonding and relay information. The advent of networked computer has inadvertently supported this notion by facilitating communication services such as electronic mail, instant messages, chat and video conferencing to be incorporated into existing computing technology. In effect, communication system that congregates users in the virtual context despite their physical proximity has been widely adopted. The emerging trend of migrating commercially successful desktop applications into mobile environment is the heart of numerous mobile researchers. Generic categorisation of the offered mobile services has been developed and classified into content, commerce and communication [1]. Hence, mobile chat is intended to be the core communication of the study in this paper.

Vronay et al. [2] has conducted an in-depth investigation on chat system and defined chat as "two to twenty or more people who appear together on a common channel of communication known as chat room." This vague explanation merely stresses on the social gathering in a public chatting space without emphasising on actual engagement of conversation and nature of communication.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 432–441, 2007.
© Springer-Verlag Berlin Heidelberg 2007

In the effort to clarify terminology, ISO/IEC JTC1 SC36 [3] has delineated chat as "a form of synchronous interactive online typewritten communication allowing users to engage in text-message conferencing. Chat is also used for private communications between a subset of participants." The definition neither associates to usability nor conceptualise nonverbal cues such as body language and facial expression that is crucial in conversation.

Evidently, a novel description that facilitates expressive communication is needed. Thus, chat is characterised as a virtual congregation of two or more participants in synchronous communication mode that enable articulate interaction via texts, graphics and images in both private and public chat rooms.

In the exertion to support mobility, the prerequisite for the design of mobile devices is the diminutive figure of the gadget. Consequently, display screen is devised to be small [4] and effective presentation of huge and diversified amount of information through the small window posed a dilemma. Furthermore, attention has been given to the processing requirement of mobile programs to ensure low utilisation of the limited processing capacity [4]. In respect of the restricted storage and memory, the similar consideration should be granted to avoid high consumption of storage and memory. Additionally, the dynamic nature of wireless connectivity introduces delay and hinders smooth interaction among the users.

Due to the physical limitation of mobile devices, maximum displayable number of turns at any one time is significantly smaller than stationary chat system which subsequently hampers usability. Contrary to immobile chat system which archives history locally, the similar situation does not apply to mobile chat system due to the inadequate storage capacity. In addition, mobile device limitations impeded mass gathering and reduced the optimal number of users logged into a chat room. Moreover, navigation via both mobile and stationary chat systems differs due to the heterogeneous input mechanisms. In the light of the variation, navigation via mobile keypad could be unwieldy and thus, requires special attention in the design of the system.

Generic problems existed between mobile and stationary chat system could be resolved by improving and enhancing the user interface of the system. In particular, chat system does not facilitate recognition of previously chatted users and the particulars related to the users [2]. Consequently, tension between users might be created as offended party viewed negatively about the forgetful party. Furthermore, individual presence of users who are not actively participating in the chat room is not conveyed appropriately [2]. Indication of the user status such as "away" and "busy" should be revealed to inform the availability of users. Other overriding theme to uncover is the inability of chat system to express complex human embodiment that is crucial in daily interaction. Thus, this paper aims to statistically evaluate text-based and graphical-based usable interface on mobile chat and draw these observations and findings to uncover driving factors that would enhance mobile chat usability.

2 Literature Review

Burak and Sharon [5] have conducted an extensive usage study on mobile services, named FriendZone. They have concluded that chat application is not suitable for mobile phones due to the small screen display and cumbersome keypad that does not

encourage fast text entry. In opposition, Grinter and Eldridge [6], [7] argued and confirmed that teenagers have flexibility in adapting to the physical restriction of the mobile devices for efficient text entry which consequently become primary motivation for the mobile chat usage. Hence, prior to the design of mobile chat system, in-depth knowledge with regards to habits and preferences of potential users should be gained to increase usability of the system and enrich user experience.

Vronay et al. [2] proposed a novel interface that is efficient to cater for high amount of users. Nonetheless, porting such design to small screen space is inappropriate as it cluttered the screen display easily. Nevertheless, we believe that conceptualising interactions in horizontal course facilitates the efficient tracking of turns in a limited screen display.

The minimalist approach adopted by Viegas and Donath [8], [9] in Chat Circles series manipulated unique colours to represent individual presence. Colours are perceived differently according to cultural and demographic background. For instance, green is perceived as unlucky in Britain and Ireland, and attire of green must be avoided in weddings [10]. Conversely, green represents longevity and life to Chinese in conjunction with their preference to green jade or nephrite [10]. Direct inference from this finding to chatting system is users whom selected biased colour to a particular community might not be welcomed and even treated harshly. Furthermore, favourable colours might be highly utilised resulting in ambiguous representation of individual identity due to the small palette of colour. By observing these differences, conclusive statement could be drawn that usages of colour for identity representation is highly controversial and critical. Viegas and Donath [9] have integrated background images which served as shared contexts to promote congregation and encouraged conversation around the images. Amin et al. [11] have supported this notion by stating that shared awareness of the settings, objects and influences in the surrounding context augment the communication and minimise misapprehension. Therefore, usage of background images should be employed as it facilitates interaction and introduces topics for discussions resulting in a usable system.

Chat system that depends on text as the sole communicative element lacks the facility to convey nonverbal cues presented in face-to-face communication. Thus, numerous investigations and experiments [11], [12], [13], [14] have conducted to enrich the user experience by integrating behavioural expressiveness such as angry and upset into avatars and emoticon. Amin et al. [11] have recommended an extension to Short Message Service (SMS) named SenseMS which allows emotion and context to be readily perceived prior to the viewing of message content to aid understanding. While this design is developed for asynchronous nature of communication, repeated display of emotional status prior to each turn in real time environment introduces redundancy and attenuates the usability of the system.

3 Research Design and Methodology

3.1 Overview

The target sample chosen was teenagers as recorded to be the forerunner of mobile chat with intention to enhance the viability of this study. There were 4 surveys

carried out in different time and locations to facilitate and coordinate the different time availability of the test subjects. The chosen local mobile chat service was Maxis SMS Chat which uses SMS as the underlying communication protocol for messages exchange.

Henceforth, Maxis SMS Chat was refereed as text-based mobile chat system (TMC) and our proposed system as graphical-based usable mobile chat system (GMC). Distinct difference between TMC and GMC was the user interface design of TMC utilised text as the sole element while GMC employed a myriad of text and graphics, and appropriate avatar and colours in the interface design as illustrated in Fig. 1. This difference could be harnessed by investigating user reception and perception on both text and graphical-based usable interfaces of mobile chat.



Fig. 1. User interfaces of TMC and GMC respectively

3.2 Demography of Test Subjects

A total of 53 test subjects involved in this experiment with demographic background that included different gender, education level and age. The test subjects were within the age range of 15 to 21 which reflected the age classification of teenagers. In all, 25 females and 28 males have participated in the survey. As a measure to effectively segregate the sample based on the education level, 27 test subjects were selected from The University of Nottingham, Malaysia, and 26 from a local secondary school in the district of Klang Valley, Malaysia.

3.3 Materials Used

There were 2 additional materials enclosed with the questionnaire distributed to the test subjects: the user documentation of Maxis SMS Chat which intended to assist the

understanding of the existing commands for service activation and utilisation; whereas the supplementary material served to define terminology found in the questionnaire that might be unfamiliar to teenagers with lower computing knowledge such as “user interface”, “usability”, “navigate” and so forth.

3.4 Experiments

The experiments for both groups took place in a control environment for 1.5 hours. The procedures began with a briefing, followed by testing and evaluation of TMC and GMC and ended with demonstration and assessment of TMC and GMC.

4 Findings, Results and Discussions

Prior to data analysis, reliability of the instrument was assessed to ensure the internal consistency holds. Generally, alpha value that exceeds 0.60 is accepted to have an internally consistent construct [15]. Both TMC and GMC constructs were internally consistent with coefficient values 0.610 and 0.616 respectively.

4.1 Hypotheses

There were 2 hypotheses formulated as initial presumptions on the relationship between user interface and usability. Table 1 details the hypotheses and the corresponding descriptions. All the hypotheses were formulated to test and evaluate at 95% (0.05) significant value as to fulfil the research objectives.

Table 1. Research hypotheses statement

$H_{\#}$	Narration
H_1	There is a significant linear relationship between user interface and usability of TMC
H_2	There is a significant linear relationship between user interface and usability of GMC

4.2 Background of Test Subjects

As coincide with the research carried out by Grinter and Eldridge [6], [7], SMS is the most preferred chatting tool among teenagers, which in this case also applied to Malaysian teenagers. Only 3 test subjects (5.7%) have not communicated via SMS while more than half (50.9%) used the service actively which were more than 20 times per week.

Multimedia Messaging Service (MMS) as anticipated and consistent with study reported by Amin et al. [11], yielding low usage as 73.6% of the sample has not used the service before. This is not surprising as the service is restricted to certain mobile devices and the cost associated to each is 2.5 times more than per SMS in Malaysia [16]. On the static Internet platform, 75.5% of the sample has used online chat service which corroborated to the findings from Pew Internet and American Life Project

[17] with a marginal difference of 0.5%. Mobile group chatting services scored an extremely low usage with a staggering 83% of the test subjects have not used the service. The low usage is expected as mobile group chatting services are not well advertised and developed in Malaysia.

4.3 Hypotheses Evaluation

Null hypotheses were statistically tested to determine the trueness of the claims. In the case of hypotheses H_1 and H_2, multiple regression analyses were conducted to test the trueness of the hypotheses with usability as the dependent variable and user interface as independent variables.

4.4 Relationship Between User Interface and Usability of TMC

Research outcomes of the multiple regression analysis indicated there was a significant relationship (p=0.003) between the user interface and usability of TMC. Based on the beta coefficient, we observed that there existed a positive linear relationship (β=0.395) between the user interface and usability of TMC in which the enhancement to the user interface that subsequently encourages usages of the service would increase the usability of TMC. This relationship could be simplified via the equation:

Usability of TMC = (User interface that encourages usage) x 0.395. [18] (1)

The beta coefficient is essential to determine the relative absolute magnitude in predicting the usability of TMC. Research findings signified that the increased of 1 unit on TMC usability requires 0.395 increased of user interface that encourages usages. Although user interface assessments are qualitative could not be effectively represented with real numbers or figures, the beta coefficient provided insight as to the magnitude of influence the variable has against the usability of TMC. Thus, the results denoted that the effort to enhance the usability on text-based mobile chat service should be placed in ensuring the user interface that encourages usages. Other variables vital to usability such as navigation, user satisfaction, and effectively were not significantly linear correlated plausibly because the importance on improving user interface for usages has overshadowed other usability aspects. Based on the analysis, H_1 experienced a rejection of null hypothesis.

4.5 Relationship Between User Interface and Usability of GMC

It is evident through the significant values that there were strong correlations among the user friendliness of interface (p=0.001), user satisfaction induced by user interface (p=0.007) and usages promoted by user interface (p=0.011) on the usability of GMC. The evaluation of the beta coefficients has highlighted that all the linear regression has positive relationship with usability.

Usability of GMC = (User friendliness of interface) x 0.402 + (Satisfaction enhanced by user interface) x 0.320 + (User interface that encourages usage) x 0.303 + (Navigation ease of user interface) x 0.171. [18] (2)

We could safely conclude that user friendliness of interface that enhances user satisfaction and encourages usages possess relatively strong magnitudes in the

correlations with usability on GMC. The user friendliness of interface has the highest absolute value (β=0.402) as compared to other variables significantly typified that it correlated strongly with the usability. The equation highlighted that 1 unit increased of GMC usability requires 0.402 increased of user friendly interface, 0.320 of satisfied user interface, and 0.303 of user interface that encouraged usage, and 0.171 of navigation ease. Evidently, work to improve the usability of GMC should be placed firstly in creating user friendly interface followed by other variables in the order of their beta coefficient values. Hence, H_2 manifested a rejection of null hypothesis.

4.6 Results and Discussions

Generally, the evaluation for each attribute in GMC surpassed that of TMC considerably with the lower mean difference being 2.42 or a staggering 48.4% rating difference between both services. These figures added the credential of graphical-based usable interface being a better design solution than text-based user interface on mobile chat in every aspect of user interface and usability that were being assessed. The findings and results supported the main distinction between text-based and graphical-based usable mobile chat was attributed to the navigation ease with GMC scoring favourably high ($\bar{x}$=4.47) as opposed to TMC ($\bar{x}$=1.45). Additionally, the user friendliness of interface ($\bar{x}$=4.51), learnability ($\bar{x}$=4.43), well integration of functions ($\bar{x}$=4.42), consistency ($\bar{x}$=4.38) and simplicity ($\bar{x}$=4.38) were the highest rated attributes for GMC evaluation. These factors are undeniably the key advantages of employing graphical-based usable design on mobile chat. In contrast, the experimental evaluation results denoted that inconsistency ($\bar{x}$=1.42), difficulty in navigation ($\bar{x}$=1.45), complexity ($\bar{x}$=1.49), the need for technical support ($\bar{x}$=1.51) and non-user friendly interface ($\bar{x}$=1.53) were the major disadvantages for text-based mobile chat. The findings significantly revealed that additional effort is required to enhance text-based user interface by improving the consistency followed by navigation, simplicity and other factors.

Principally, preferences between TMC and GMC were consistent among all demographic background variables. Only minor difference tabulated in evaluating the user interface and usability for both services. The assessments between usability and user interface for the same service only differed marginally with mean value that was less than 0.2. The test subjects in the sample have assessed GMC favourably in both aspects of usability ($\bar{x}$=4.32) and user interface ($\bar{x}$=4.37). On the contrary, the usability ($\bar{x}$=1.62) and user interface ($\bar{x}$=1.60) of the text-based mobile chat service was not gaining much support from the test subjects. Evidently, the research results elucidated that graphical-based usable mobile chats were perceived to have highly valued user interface which contributed to the usability of the service; whilst text-based user interface was not well appreciated to be a usable mobile chat.

The implications on mobile chat usability are noticeable. Research outcomes from the empirical evaluation were positive. In addition, the optimistic results typified the

importance aspects and benefits that we could gain from employing a graphical-based usable interface on mobile chat. Conversely, in the effort to enhance the usability of text-based mobile chat, primary attention should be devoted in creating a user interface that promotes usages. This could be achieved by firstly understand the motives for the usage such as effectiveness, efficiency or satisfaction induced by the service. Although, no other significant relationship could be drawn to relate to the usability of the TMC, variables such as navigation ease and consistency of interface should not be quickly ruled out as trivial. This is due to the interpretation that upon invoking usage desire via the user interface, utilisation of the service might increase and subsequently, other variables should be given sufficient considerations.

As for the graphical-based usable mobile chat, usability could be augmented by proving user friendly interface that enhances user satisfaction, promotes usages and navigation ease. The findings significantly provided sufficient information for mobile chat designers to focus in enhancing the user interface and usability of mobile chat. Specifically, the needs of the target chatters should be identified to act as a framework for designing highly usable mobile chat that would consequently promote user satisfaction, user friendliness, usages and navigation ease.

5 Conclusion

Current mobile chat systems fail to replicate the success enjoyed by the static counterpart as the user interface design lacks in considering the mobile devices limitations. In particular, navigating through long chat archive in restricted screen display is cumbersome and time consuming. The effort to track a specific message sent by chatter includes scrolling through the archive and identifying each message in turn. The lack of usability in existing system motivated the needs to explore a novel interface to enhance the system effectiveness, efficiency and user satisfaction.

Although existing literature [5] cast pessimistic view on the future of mobile chat systems as the platform are deemed as not suitable for group communication, and novel interface that work around the device limitations could be a solution to promote mobile chat usages. The empirical evaluation outcomes corroborated to the research objectives and highlighted that the main difference between both services were the navigation ease with GMC leading significantly. Moreover, the research findings concluded that the user interface and usability of GMC were very well received unlike TMC which was perceived below average. On the contrary, TMC usability could be improved by creating user interface that encourages usages; whereas GMC usability is augmented by crafting user friendly interface that enhances user satisfaction, promotes usages and navigation ease. The used of specific components such as avatars, graphics and colours that constructed the graphical-based usable interface have drastically well addressed and improved the user satisfaction and navigation, and indirectly typified a strong linear relationship between user interface and usability of mobile chat. Additionally, the experimental results uncovered the major driving factors which include the user friendliness of interface that enhances user satisfaction, encourages usages and navigation ease extensively hold strong magnitude and correlated with the usability on GMC.

In short, text-based mobile chat pales in comparison to that of the graphical-based counterpart. Moreover, every aspect of usability and user interface of the graphical-based usable mobile chat is well appreciated by the test subjects. The empirical evaluation results have highlighted the potential use of graphical usable mobile chat as a substitution to the text-based chat that has poor reception and is under utilised in the commercial arena.

References

1. Chae, M., Kim, J.: What's so different about the mobile Internet? Communications of the ACM, pp. 240–247 (2003)
2. Vronay, D., Smith, M., Drucker, S.: Alternative interfaces for chat. In: Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology, North Carolina, USA, pp. 19–26 (1999)
3. ISO/IEC JTC1 SC37.: Abstract collaborative workplace conceptual architecture contribution. [Online] (2004) Accessed on 18th November 2005 Available: WWW URL http://collab-tech.jtc1sc36.org/doc/SC36_WG2_N0077.pdf
4. Paelka, V., Reimann, C., Rosenbach, W.: A visualisation design repository for mobile devices. In: Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction, Cap Town, South Africa, pp. 57–62 (2003)
5. Burak, A., Sharon, T.: Usage patterns of FriendZone – mobile location-based community services. In: Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia, Maryland, USA, pp. 93–100 (2004)
6. Grinter, R.E., Eldridge, M.A.: Y do tngrs luv 2 txt msg? Why "texting" became popular with teenagers. In: Proceedings of the 7th European Conference on Computer Supported Cooperative Work (ECSCW), Bonn, Germany, Kluwer, pp. 219–238 (2001)
7. Grinter, R.E., Eldridge, M.: Wan2tlk?: Everyday text messaging. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems, Florida, USA, pp. 441–448 (2003)
8. Viegas, F., Donath, J.: Chat circles. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems: the CHI is the limit, Pennsylvania, USA, pp. 9–16 (1999)
9. Viegas, F., Donath, J.: Chat circles series: Explorations in designing abstract graphical communication interfaces. In: Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, London, England, pp. 359–369 (2002)
10. Hutchings, J.: Colour in Folklore and Tradition – The Principles. Colour Research and Application 29(1), 57–66 (2003)
11. Amin, A.K., Kersten, B.T.A., Kulyk, O.A., Pelgrim, P.H., Wang, C.M., Markopoulos, P.: SenseMS: A User-centered approach to enrich the messaging experience for teens by non-verbal means. In: Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services, Salzburg, Austria, pp. 161–166 (2005)
12. Berg, S., Taylor, A.S., Harper, R.: Mobile phones for the next generation: Device designs for teenagers. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florida, USA, pp. 433-440 (2003)
13. Persson, P.: ExMS: An animated and avatar-based messaging system for expressive peer communication. In: Proceedings of International ACM SIGGROUP Conference on Supporting Group Work, Florida, USA, pp. 31–39 (2003)

14. Smith, M.A., Farnham, S.D., Drucker, S.M.: The social life of small graphical chat spaces. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems, The Hague, The Netherlands, pp. 462–469 (2000)
15. George, D., Mallery, P.: SPSS for Windows step by step: A simple guide and reference 10.0 update 3rd edn., Allyn & Bacon, Pearson Education Company, USA (2001)
16. Maxis Mobile.: Maxis SMS chat. [Online] (2002) Accessed on 23rd February 2006 Available: WWW URL http://store.maxis.com.my/pdf/SMSChat.pdf
17. Fox, S., Madden, M.: Generations online December 2005. [Online] (2005) Accessed on 27th March 2006 Available: WWW URL http://www.pewinternet.org/pdfs/PIP_Generations_Memo.pdf
18. Lacey, M.: Multiple linear regression. [Online] (1997), Accessed on 23rd March 2006 Available: WWW URL http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm

A Tale of Two Teams: Success and Failure in Virtual Team Meetings

Marilyn M. Tremaine[1], Allen Milewski[2], Richard Egan[1], and Suling Zhang[1]

[1] New Jersey Institute of Technology, U.S.A
{tremaine,egan,sz29}@njit.edu
[2] Monmouth University, U.S.A
amilewsk@monmouth.edu

Abstract. Interaction between two teams with the same team leader and with similar size and goals moved from weekly face-to-face meetings to virtual meetings because of the temporary displacement of the team leader to a time zone six hours ahead of the rest of the team. One team focused primarily on software development and the second team on developing and testing a research instrument. The Software Team floundered through multiple different meeting arrangements and eventually agreed to disperse until the leader returned to the same time zone. In contrast, the Research Instrument Team kept a single meeting time that was set before it moved to virtual gatherings, and continued to be an active and productive team. This paper explores what factors led to this divergence in team success and concludes that the implicit temporal structures entraining the members of the Software Team coupled with an inability to repair member unhappiness and an unequal dispersion of skill sets among virtual and co-located members led to one team's eventual shutdown.

1 Introduction

Global teams are a critical part of the workforce in both software and usability engineering [1,3,4]. They have high potential to gain time-to-market advantages by use of a twenty-four hour work cycle. In addition, wage advantages can be had by outsourcing work to countries with lower costs. Finally, emerging markets make it advantageous for companies to distribute their labor forces to countries where new markets are perceived [6].

Even with the spread of the internet, advances in development and management tools, increased education standards and increases in English as the de facto technology language, global teams have not worked as well as hoped [8, 11]. Productivity has often been low, and even successful teams are fragile and plagued with multiple problems. Some of these problems are postulated to stem from cultural differences; others from the physical and temporal distances that limit communications; still others from the limitations of the communication medium used to manage teams.

This work investigates two design and development teams: one that smoothly became a successful virtual team and the other whose transition to "virtualness" experienced serious difficulties. Prior to becoming virtual, both teams met regularly

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 442–451, 2007.
© Springer-Verlag Berlin Heidelberg 2007

face-to-face, and both teams were highly productive. Both teams had the same team leader and were the same size. The teams moved to a virtual setting when their team leader temporarily relocated to another continent for six months. Because the team leader attended meetings virtually, all members of both teams used this opportunity to also participate in the meetings virtually. Within three months, one team continued to meet on a regular basis and be highly productive. The other team struggled with meeting attendance for eight months has now been closed down until the team leader returns.

We explore the causes for the divergence in these teams through a series of one-on-one telephone interviews and through a review of the detailed meeting minutes. In the course of the interviews, many factors were identified as potential reasons for the divergence in the two teams, but two factors were found to be critical differences. First, we found that implicit temporal structures entrained the unsuccessful team to the extent that members were not able to find a suitable time in which all members could effectively attend meetings [12]. These temporal entrainments did not, in and of themselves, lead to the slowdown in the team. Rather, they instigated a series of seemingly minor problems such as, one or another party might be kept waiting for a virtual meeting for long periods without knowledge of what was causing the delays. This led to tension amongst team members, and this tension remained unrepaired for long periods of time. The key effect of virtualness in this case was the lack of effective ways to perform the needed repair. [11]

The second factor was interpreted to be an interaction of leadership style and skill distribution. On the successful team, the distribution of skills was somewhat even; all members had diverse skills and the team roles could be interchanged. In the other team, there were greater differences in the types of skills members possessed, and as a result, the team roles were much stricter. This was especially acute because it was the remote team leader who possessed the most unique skill, that of team management and interface design. In addition, the location of software development skills in the three co-located members led to a formation of a sub-team that created its own agenda and worked separately from the rest of the team. The casual management style of the leader allowed this sub-team to form and also to set its own design agendas. Similar results on subteams have been reported elsewhere [4].

In summary, the divergence of two seemingly similar teams with a common leader has provided a natural study permitting us to draw comparative insights about factors that contribute to global teams' success or failure. We explain a variety of small problems that led to one team's shutdown and describe a set of relatively trivial procedures that the struggling team could have attempted to circumvent their problems.

2 Description of the Software Team

The Software Team has been in existence for about four years. The team is involved with all aspects of the software product being produced, from conducting exhibits at trade shows, conducting user studies, negotiating contracts with other software vendors and writing descriptions of the software product being developed.

A key feature of this team is that all of the work being done by the team members is voluntary. The team is producing a downloadable software product to help the blind

and visually impaired. The product reads RSS feeds from online newspapers and other internet sites, converting the information from text to speech. It is also built using open-source software with the intent of making the downloadable package free to users. The team's goal during the virtual year was a release its first software version and the design of an additional application for the interface.

The team has five members and has had fifty per cent turnover since its inception. Three of the members are computer scientists who manage the coding, maintenance and testing of the software product. A fourth member is the team leader and a fifth member is the usability engineer who also has web development experience. The team has regularly met weekly for two hours at 6 PM on Wednesdays including socializing time and dinner. This team could be described as a viable, lively and highly interactive team. Between meetings, there was significant email exchange, and team members were constantly finding out new web sites, papers or products that other members were encouraged to look at [10]. Agendas for each meeting were solicited ahead of time. Issues were brought up and discussed and consensus was invariable achieved. Often the software team would arrange a second UML design meeting or debug meeting to resolve some of their issues. In the last year before the team went virtual, the group completely rewrote the software product in Java, added a new text-to-speech engine, changed the input mechanism, worked with a set of blind consultants to obtain feedback on the designs and built a robust enough software platform to exhibit the product at two trade shows.

The team used a Yahoo™ Group to manage itself. A calendar of meetings with automatic reminders was kept on the Yahoo Group and team members used the Yahoo ListServ to send emails to the entire team and uploaded papers and documents to the Yahoo Group.

3 Description of the Research Instrument Team

The team has four members. Two of the members are working on their doctorate and the instrument development is part of their research. The other two members are faculty members interested in the research project. The team leader is the thesis advisor for the two students. The key reason that the team is a joint team is the mutual interest in global software development.

The goal of the Research Instrument Team has been to conduct a survey to be distributed to a set of Fortune 500 companies on issues relating to global virtual team management. The work involved developing and conducting the survey, analyzing results, and setting up additional corporate relationships. This team had been in existence for about a year before it went completely virtual. The team met once a week during the year to discuss issues and plans. One of the members frequently attended the meeting by telephone throughout the year because of time commitments elsewhere. When long range planning had to be done, the team met for an all day retreat on Saturday.

The team had generated two papers and two doctoral consortium proposals before it went virtual. It also developed, validated and tested the reliability of two survey instruments that capture and model what are believed to be relevant factors affecting

team performance and satisfaction. The survey instruments were piloted in four separate team-based courses at two universities.

Face-to-face meetings typically lasted two hours. Much discussion took place over the research issues that the team faced. In addition, many research hurdles associated with partially completed surveys, incomplete team responses and subject scheduling had to be dealt with. Much of the activity of the team members during the time between meetings was spent ferreting out information that was needed or negotiating survey schedules with course instructors. Meetings were focused on problem solving and there was less socializing and off subject conversations than in the other team.

This team also uses a Yahoo Group to manage itself. The Yahoo Group was primarily used for uploading papers to be shared with other team members. The mail service part of the Yahoo Group was not used as frequently, and documents created by the group tended to be exchanged via email rather than through the Yahoo Group repository.

The Research Instrument Team had less social exchange than the Software Team. Email exchanges usually occurred a day or so before each meeting as meeting members shared the work they had done during the week.

4 Management Practices of the Team Leader

Both teams became virtual in late June of 2006. They have been meeting or attempting to meet for the months of July through January. The team leader of both teams implemented the following management practices for both teams in an effort to counteract some of the problems the teams were likely to experience with the limited bandwidth of the communication tools that were selected.

Both teams connected using a voice over IP conferencing system. One person in each group became the designated conference originator. Initially, all attendees were virtual from both teams.

The following procedures were followed for both meetings:

- An agenda was established for the meeting
- Action items from the previous meeting were reviewed in the agenda.
- The team leader took the minutes for the meeting
- The team leader also served as the facilitator of the meeting.
- Documents that had been mailed to team members or posted on the team's group were discussed
- A social exchange was scheduled for the end of the meeting in which each team member was asked to tell something fun that had been done during the week.
- Minutes of each meeting were sent to all team members right after the meeting

These practices deviated from the face-to-face meetings. In particular, the Research Instrument Team was less likely to use an agenda in face-to-face meetings. Neither team maintained minutes when the meetings were face-to-face, although individuals took notes during the meetings.

5 Meeting Difficulties for Both Teams

Initially, there were significant difficulties with the low cost digital conferencing technology being used by both teams. One of the key issues that team members faced was degradation in voice quality and connection problems that depended on the available instantaneous bandwidth of each individual's internet service. Speech was frequently broken up and the speaker was asked to repeat a statement – sometimes three or four times.

> *"Its pretty powerful, but the downside is that it is still a bit of a fragile environment. We're relying on networks….being up…..with decent throughput…that everyone has the same version of the software…and that peoples PCs work."* (Software Team member)

Another difficulty with the meeting was background noise. This problem was resolved over time as team members purchased headsets or a standalone microphone to avoid this problem. The Software Team ran tests on the quality of sound with different options until they arrived at a usable solution. Their solutions were used to help the Research Instrument Team.

A larger meeting difficulty was the difficulty of walking through very complex material remotely. The Software Team walked through UML diagrams and the Research Instrument Team discussed reams of statistical analysis runs. If comprehension became too difficult at one meeting, the responsible team member was given an action item to prepare a summary document that would make the discussion clearer at the next meeting.

Team members at both meetings indicated that team meetings could be tedious and that they would often catch themselves reading email, cleaning up their filing system or playing computer games.

6 Meeting Difficulties for the Software Team

One of the key problems encountered by the Software Team was the inability to come up with a viable meeting time. Meetings were set to begin at 6 PM, with the leader starting the meeting at midnight since she was six hours ahead. Since the meeting lasted approximately three hours, including pizza time, this meant that the team leader would be heading home at 3:00 AM.

The next meeting attempt was to set the meeting at noon. This would be 6:00 PM in the team leader's country. Because of limited lunch hours, the team could not meet for more than one hour. Additionally, two of the attendees were often called to other meetings on short notices.

The third attempt was to schedule the meeting at 7:00 AM for four of the team members and 1:00 PM for the team leader.. One team member never came to this meeting and two others missed this meeting once or twice because they could not get up at 7:00 AM.

To meet the time zone constraints of individual members, it was decided as a fourth attempt, to hold the meetings on Saturday at noon. The first virtual meeting

was tried. One of the team members forgot about the meeting and a second member decided not to show up.

A fifth solution was tried, that of making the meetings into face-to-face meetings except for the distant team leader. Meetings were set to occur on Saturdays. However, attendance at the meetings was still poor. In addition, meeting members would come to these meetings late making it unclear to the on-time members whether a meeting could take place or not. Since all team members had to travel some amount to make it to the meeting and since meetings scheduled on Saturday took time from other scheduled activities, committing to a meeting that then did not occur became a significant deterrent to attending the next meeting.

A sixth solution presented itself with the arrival of home internet by the Team Leader, and 6:00 PM (midnight) meetings were set for Wednesday each week. This was the team's normal meeting time when they were all co-located. A key problem with this arrangement was that the team members often gave notice that they could not attend a meeting approximately one hour before the meeting. Many of the reasons for non-attendance were, "going out with a friend," "not feeling too well," "have a family event to attend," etc. For those members who did make the meetings, there began to be a buildup of frustration with what was perceived as a cavalier attitude of the other members. In mid-January, the two exasperated managers agreed to stop having team meetings until further notice.

The problem with each of the attempted meetings was the loss of one to three members. In a team as small as the Software Team, this slowed work considerably, so the team could not make progress on important issues until that member again joined a meeting. Eventually, the forward momentum of the project stopped and the team meetings involved going over the same material multiple times. Meetings became boring rather than the rapid exchange of new ideas.

> *"Things fell apart from a communication standpoint…. There was just misunderstandings like that. You send an email and you assume that somebody has read it whether somebody has read it or not. There was a situation where an email was sent out to confirm a meeting and several group members assumed that not replying was implicit agreement whereas when the meeting came about and we had [technical] difficulty getting connected, the [remote team leader] didn't see the team come on in ten minutes and then assumed that no one was there. And, because of the lack of email [further] assumed that no one was participating."* (Software Team member)

7 What Caused the Software Team to Fail?

We have presented two successful face-to-face teams in which members were highly productive and very satisfied with their teams. While the Research Instrument Team was somewhat more culturally diverse than the Software Team with one member coming from another country, the team composition was quite similar in the two teams. Both teams had a wide age distribution.

Possible factors in the Software Team's slowdown are listed and addressed in the following paragraphs. Analysis of these factors points to two that appear to be causally critical.

- *The Software Team had less motivation overall, or less "momentum" because, e.g. their team's work happened to be at a natural stopping point when the team became virtual* -- We considered differences in initial motivation between the teams, but participation in both teams was voluntary, making us discard this possibility. All team members did gain authorship on papers coming from their work. While Software Team members did not need this benefit because they worked as programmers in the industry, it seems unlikely that this would cause a serious difference.
- *The Software Team met primarily for social reasons and lost this socializing when the team became virtual* -- There are two reasons why the loss of socializing through a team going virtual does not apply. First, large parts of the software development end of the Software Team already worked virtually. They would often share screens and use chat to solve nasty problems that came up. A social exchange was also tried at the end of the Software Team's meeting, i.e., members were asked how they had spent their weekend, but team members were not as eager to participate in this social exchange as the Research Instrument Team was.
- *The Research Instrument Team really had two managers, that is, two faculty members, with one member still being local and able to maintain the momentum of the team* -- Although the Research Instrument Team ostensibly had two managers, the Software Team also had a local and remote team leader. One of the key benefits that the software development end of the team obtained from team membership was mentoring. One of the Ph.D. students on this team was a senior level software developer at a prominent computer company. He was in charge of the software development for the team. The two other team members continually learned software tricks from him which enhanced their job skills significantly. Much of the between meeting exchange in the Software Team was in solving software problems and working with the team's software manager.
- *The Software Team was carrying out a unique task - software development. This task has different properties that are not suited to virtual teams* -- It can readily be argued that there are many similarities between software and survey instrument development. For instance, in both cases, work is product-oriented – developing software and survey. In both cases, the pace of the work was driven by self- and team-imposed deadlines. The detail that has to go into question design and the order of questionnaire layout is similar to that required in software. Walking through a set of thirty regression analyses to determine an overall model of effects is as complicated as debugging code. We also observed similar mentoring between more skilled and less skilled team members but not at the level that occurred in the Software Team.
- *The Software Team could not find a convenient time to meet* – The scheduling issue is perhaps the most compelling cause of difficulties, and is acknowledged by team members. While scheduling is a likely factor, its effect is indirect. It appears that the key factor is actually frustrated expectations resulting from scheduling accommodations.

"I think the majority of those problems was people having their scheduling conflicts and then put , you know, the six time zones and the communications failures on top of that…things just went kind of ugly for probably a month, month and a half" (Software Team member)

Because all business was transacted at the main meeting (unlike the Research Instrument Team which also had many individual meetings) there was no way for individuals to express their dissatisfaction with anyone else's team attendance behavior without making it a meeting wide issue. None of this discontent was expressed in the conference calls. Apologies were made for late arrivals, but not for last minute decisions not to show up for meetings. Meetings became discussions of what to talk about at the next meeting when everyone shows up. The priority for attending the Software Team meetings had dropped in each member's ratings so that other events took higher priority, including getting sleep.

- *The Software Team's distribution of skills was uneven and distributed geographically* -- A second reason for the team's demise was the tacit breakup of the team into two subteams.

  First, a Senior Usability Leader left the project almost as it went virtual because of consulting commitments. The Team Leader's expertise was also human-computer interaction. This created a team where the virtual part was creating the user interface design and the co-located part was developing the software for the design. The co-located group began to make more design decisions and the meeting discussions became more about software issues than interface design issues. The Team Leader was not happy with some of the design decisions being made but felt that it was better to let the Software Team run with the decisions in order to keep up the momentum of the project. Both user interface people began to feel irrelevant to the project as the document exchanges and meeting conversations became more technical. The Usability Engineer commented, *"I don't attend because there is nothing for me to do."* In interviews with the software team, they expressed this distinction as a preference to discuss code related issues during meetings.

  The software team was rather surprised when the team leader worried about the team failing. They felt that they were accomplishing a lot and working quite hard on the project. They commented, *"yes, there were some time difficulties in setting up meetings but they were not serious."* They did not feel that the issue with missed meetings and lateness was at all serious.

  The local software team also did not feel the need for the weekly meetings because they were in constant contact through phone calls, email, a version control software program and instant messages. Yet, the weekly meetings were the primary contact of the usability and design side of the group [14].

A rudimentary version of the software was made available for download on the group's website. In January, the Team Leader and Software Manager decided to end meetings for the moment because it was too much effort to keep them running.

8 Conclusion

In the above discussion, we have carefully reviewed a series of reasons for why one team had troubles with virtual meetings and a second team managed to continue to perform productively. Interviews with members of the team having troubles indicate underlying issues that were not a part of the successful team. A buildup of discontent occurred with the team leaders and the other team members that was not brought out and discussed with team members. When it finally did come out, through the interviews, some team members were surprised that the team leader thought problems existed. Members who had regularly missed or had been late for meetings did not consider this behavior to be an issue. In addition, none of the co-located members felt that they were deviating from the goals of the project. In contrast, the team leader felt that step-by-step, the original design had been eroded so that what was to be made available no longer contained the usability characteristics intended. Unfortunately, the virtuality of contact made it hard for her to convey some of these design issues. In addition, there was also a tendency to argue against her ideas for practical reasons from the tightly knit co-located part of the team. Because the team consisted solely of volunteers, the leader was also concerned that imposing too many restrictions on the work of the team would lower their motivation and potentially lose them as team members. In the end, both the team leader and software manager were frustrated with other team members temporal reliability and the direction the project was heading.

These results suggest that temporal constraints indirectly affect virtual team performance as does the distribution of team member skill sets. Thus virtual team management needs to look at these issues when setting up virtual teams. But more important, could something have been done that would have prevented the problems with the Software Team. The answer is unequivocally, yes. Below, we list a variety of activities that would have prevented the Software Team's demise.

- When the co-located usability people left the team, they should have been replaced.
- The anger and frustration with the missed meetings and the late arrivals should have been discussed in individual meetings with the team leader .
- Meetings in which member goals are stated and differences worked out should have been held at regular intervals.
- Meeting tools which allowed for richer presentation of difficult concepts should have been regularly used.

These are simple fixes. They take work but can be put in place. The current reason they are not being put in place is because the virtualness of the team was, by plan, only temporary as the team leader will be returning the United States.

The key point in this paper is that very trivial items caused very large problems in a virtual team that was neither that far apart in terms of time zone differences or in terms or team member differences. Outsourcing and off-shoring to globally constructed teams which can be expected to be much further apart than the team discussed are certain to suffer from an exacerbation of the above two problems.

> *"I listed probably what I think were about ten factors…and all of them are trivial by themselves, but in aggregate form, where it becomes like the perfect storm of the group not working well* " (Software Team member)

In Memoriam

This paper is in memory of John Visicaro, a member of the Software Team who suddenly took ill and passed away on January 21, 2007. John was only 43 when he died and was such a vibrant and important member of the Software Team that he will be sorely missed.

References

1. Aykin, N.: Usability and internationalization of information technology. Lawrence Erlbaum, New York (2005)
2. Bluedorn, A.C., Denhardt, R.B.: Time and Organization. Journal of Management 14, 299–320 (1988)
3. Borchers, G.: The software engineering impacts of cultural factors on multi-cultural software development teams. 25th international Conference on Software Engineering, Portland, OR USA (2003)
4. Bos, N., Shami, N.S., Olson, J.S., Cheshin, A., Nan, N.: In-group/Out-group Effects in Distributed Teams: An Experimental Simulation. In: Proceedings of Computer Supported Cooperative Work. CSCW'04, ACM Press, New York (2004)
5. Carmel, E.: Global Software Teams Collaborating Across Borders and Time Zones. Prentice Hall, New York (1999)
6. Cramton, C.D.: Attribution in distributed work groups. In: Hinds, P., Kiesler, S. (eds.) Distributed Work, The MIT Press, Cambridge, MA (2002)
7. Curtis, B., Krasner, H., Iscoe, N.: A Field Study of the Software Design Process for Large Systems. Communications of the ACM 31(11), 1268–1287 (1988)
8. Damien, D., Zowghi, D.: An insight into the interplay between culture, conflict and distance in globally distributed requirements negotiations, presented at 36th Hawaii International Conference on System Sciences (HICSS'03), HI, USA (2003)
9. Herbsleb, J.D., Moitra, D.: Global Software Development. IEEE Software 18, 16–20 (2001)
10. Jarvenpaa, S.L., Leidner, D.E.: Communication and Trust in Global Virtual Teams. Organizational Science 10, 791–815 (1999)
11. Milewski, A.E., Mullinix, B.: Solitary Collaboration Across Cultures: A Qualitative Examination. In: Aykin, N., Preece, J. (eds.) Human Computer Interaction- International, Lawrence Erlbaum, Mahwah (2005)
12. Olson, J.S., Olson, G.M.: Culture Surprises in Remote Software Development Teams. Queue 1, 52–59 (2003)
13. Orlikowski, W.J., Yates, J.: It's about time: Temporal structuring organizations. Organizational Science 13(6), 684–700 (2002)
14. Poltrock, S.E., Grudin, J.: Organizational obstacles to interface design and development: Two participant-observer studies. ACM Transactions on Computer-Human Interaction 1, 52–80 (1994)

Assumptions Considered Harmful
The Need to Redefine Usability

Heike Winschiers and Jens Fendler

Department of Software Engineering
Polytechnic of Namibia
13 Storch Street, Private Bag 13388
Windhoek, Namibia
{heikew,jfendler}@polytechnic.edu.na

Abstract. A cultural evaluation of Usability Engineering in the Namibian context reveals a number of good practices as well as locally inadequate methods. One major challenge in cross-cultural Usability Engineering is the implicit western understanding of usability and its associated assumptions which often lead to a locally inappropriate usability evaluation. Conceptualisation sessions held with different Namibian user groups confirmed a deviating perception of the term "usability". None of the groups mentioned terms "commonly" associated with "usability" such as speed, learnable, or memorable. Thus standard usability testing comprises a dual bias through the western definition of usability and the related choice of methods which aim to test an already biased objective. We therefore suggest an ethno-centric software development framework which incorporates a contextual redefinition of usability.

1 Introduction

Usability Engineering (UE) has become an increasingly important aspect of local and international software development. Software engineering paradigms like user-centred and agile development, interaction and participatory design, established the relevance of user concerns and so-called user-friendly systems. However UE, considered a subset of development processes, is often reduced to a few distinct activities and is expected to deliver specific results. Moreover, the basic idea of tailoring software to be effectively and efficiently used by a specific group of end-users has so far only been based on assumptions and experiences from the western culture's point of view. Thus in the early days internationalisation encompassed only the customisation of a fully developed application to national requirements such as language, measurements and other units. Further research indicated a deeper relation between culture and User Interfaces and system usability. However "*a major impediment in global user interface development is that there is inadequate empirical evidence for the effects of culture in the UE methods used for developing these global user interfaces*" [1].

'*We should recognise another inherent limitation of UE, that is it provides a means of satisfying usability specifications and not necessarily usability*" [2]. Thus

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 452–461, 2007.
© Springer-Verlag Berlin Heidelberg 2007

more attention should be paid to the validity of the specification. Especially in a cross-cultural setting it seems that the discrepancy between the specification and the understanding of usability is high and often leads to the development of unusable systems.

In this paper we provide empirical support for a cultural adaptation of UE methods and processes based on a Namibian case study. We further suggest a software development framework which incorporates a contextual redefinition of usability to extend current internationalisation efforts.

2 Internationalisation Efforts

Much research has been done in internationalisation of software, yet this is mostly aimed at a first level: fast product adaptations rather than analysing underlying development processes. Thus in the 1990's localisation efforts merely concentrated on national customisation. What has become known as Interna-

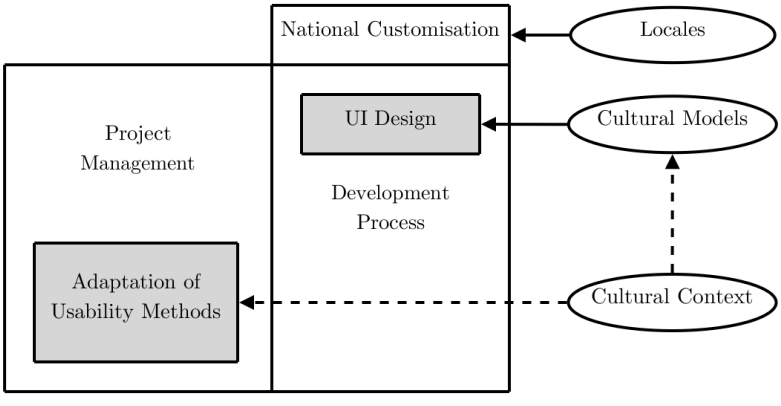


Fig. 1. Current Internationalisation Efforts

tionalisation and Localisation is often a marketing tool rather than an attempt to narrow the cultural gap between an application's potential user groups. It was soon realised that this was insufficient to gain local acceptance. Del Galdo and Nielsen [3] suggested two additional levels for consideration: the adaptation of usability methods to specific countries and the design of user interfaces to fit cultural models of how people work and communicate. In figure 1 we summarise current internationalisation efforts as described above.

2.1 Cultural Models and User Interface Design

It is widely recognised that User Interface features carry cultural values. The way signs, symbols, and colours are interpreted differs from culture to culture. How users relate to navigational structures and classification systems depends on the

way their society organises and models the world. Skills and associated assumptions about reading and sources which have been built up over a lifetime often prevail. Web site design is thus inconsiderate toward oral cultures. Walton and Vukovic have demonstrated a cultural dimension for web-information-seeking practices [4]. Among other findings they observed South African students unable to operate tree structures and breadcrumbs, concepts unfamiliar to their culture. They conclude that "*in developing contexts, the user's goals and practices may be vastly different from our assumptions, and they may not be able to crack the many codes by which we have encoded the scent*" [4].

Trillo points out that developers need a methodology to select an appropriate cultural model to guide the international user interface design [5]. Among the most popular has been Hofstede's cultural model [6] in which he distinguishes cultures along the following dimensions: Power-distance, Collectivism vs. Individualism, Femininity vs. Masculinity, Uncertainty avoidance, and Long- vs. Short-term orientation. Marcus and Gould illustrate the inference of those dimensions and User Interface characteristics with a set of selected Web sites [7]. Furthermore they attempt to derive general guidelines for user-interface and Web design, e.g. the level of power distance should be aligned with the information structure, use of hierarchies and security features. Anticipated derivations of cultural dimensions into specific user interface design rules often lead to inadequate generalisations. Ford and Gelderblom found no correlations between South African users' performance and the use of websites displaying dimension-specific characteristics [8]. Fitzgerald concludes that cultural dimension models seem to be aimed at a description of culture rather than as a prescription for best user interface design [9]. Yet we consider cultural models to be valuable sources of information for the local appropriation of usability methods. Similar opinions have been raised in the literature, where Hofstede's cultural dimensions model is suggested to inform the selection of usability assessment techniques in cross-cultural user testing [1].

2.2 Cultural Validity of Usability Methods

The success of a method depends in part on its compatibility with the context of application. There are many accounts of usability engineers having crossed cultural borders and encountered unexpected situations which can be found in the literature. E.g., the usability expert who was flown in from New York to Tokyo and would not understand why the females in his focus group were not participating [5]. The choice of method was obviously inappropriate. We have had similar experiences especially with the use of questionnaires as a valid data gathering method. Most Namibian nationals fill in questionnaires with the assumed expected answers rather than the personal truth [10]. In oral communications similar conventions can be observed. We believe that it is motivated by a cultural listener-satisfaction and conflict avoidance habit. An ingenuous interaction is further hampered by an unusual high power distance. Vatrapu

and Pérez-Quiñones experienced that when the usability expert and the user are from a different culture, usability problems may be masked rather than uncovered within a structured interview session [1]. Cultural influences on the use and success of well established methods, such as Think-Aloud task analysis and metaphors have also been reported [1].

Consequently, once the cultural determinants are known, the methods can be adapted or entirely different methods can be chosen. For example, Vatrapu and Pérez-Quiñones suggest that interviewers from the same culture might be more effective in eliciting usability problems especially when users come from hierarchical cultures [1]. In terms of Hofstede's dimensions Namibia can be characterised as a high power-distance and a rather high collectivistic culture. Elders have to be respected; this includes strict obedience towards parents, teachers and bosses. Many grass roots projects have failed because the village elders were not involved and to get employee participation employers will have to support or even order it. In terms of collectivism, large family bonds are in place supporting individual members but also demanding responsibilities. Outcomes obtained in our evaluations confirm results reported in the literature, for e.g. better qualitative feedback in interviews are obtained if usability evaluators and users belong to the same ethnic groups [11]. We have further successfully introduced a collective usability evaluation method in the form of workshops rather than individual user evaluations to reflect local community habits. Similarly can African traditional story-telling be mirrored to design task-analysis evaluations as it creates the necessary contextuality for users to relate to a task.

2.3 Invalid Assumptions – A Cross-Cultural Challenge

Besides the need for a cultural adaptation of usability evaluation methods major discrepancies between our and the users' assumptions in regard to the concept of Usability became evident. For example, most users did not complete their tasks however they felt they had mastered the system quickly and easily, and were therefore satisfied with the system. The widely presumed correlation between user satisfaction and efficient and effective task completion does not hold in the Namibian context. We were therefore surprised when we observed Namibian participants evaluating information systems by measuring the system content against their own knowledge and once they discovered the system lacks information they lost trust in the system and rejected using it [11].

Thus a consideration of the semantics of "Usability" in each context of use becomes a necessity. Allen and Buie have looked how different frequently used terms in UE, such as intuitive, user-friendly, logical, tester could be compromised [12]. They conclude that the terms must be used with care in order to hold their value and ensure a common meaning among the concerned group otherwise they can create a reality that is different from the one intended. Moreover if the group consists of people with different cultural backgrounds a mutual understanding needs to be explicitly established.

3 What Really Does "Usability" Mean?

Only few authors explicitly have defined "usability", thereby contributing to the establishment of a common assumption of its meaning. Yet concepts are hardly static entities. They evolve over time with their properties and meanings being subject to change. Over time, any concept's enclosing context evolves and in turn influences the definition of the concept itself.

3.1 The Origin of "Usability"

Usability engineering is rooted in the modernist or enlightenment tradition which values rationalism, individualism, information, performance and efficiency anchored in the definitions and measurements of "usability". According to Shneiderman "usability" can be quantified in terms of time to learn, speed of performance, error rate, retention over time and subjective satisfaction [13]. Shneiderman refers to an early US military standard MIL-STD-1472 for human engineering design criteria, in which the achievement of effectiveness, simplicity, efficiency, reliability, and safety of system operation, training, and maintenance is spelt out [14]. Similarly does Dix refer to effectiveness, efficiency and satisfaction [2]. Preece breaks usability down into the following goals: effectiveness, efficiency, safety, utility, learnability and memorability [15]. Leaving the central definition of usability untouched, Preece complements it with user experience goals, such as satisfying, enjoyable, motivating. She considers usability goals to be central to interaction design and operationalised through specific criteria while user experience goals to be less clearly defined. Most other definitions found in the literature either refer directly or indirectly to the above mentioned or just rephrase usability to be "ease of use" or "user-friendly" which does not contribute to the understanding of the term. Thus most researchers and practitioners do not question or attempt to widen the concept itself but focus on the evaluation methods. However these methods are implicitly linked to the perceived understanding of the concept. Industry-recognised methods for evaluating a system's usability, such as GOMS, focus on efficient and accurate performance [16]. Task-analysis methods intend to measure the effectiveness of the user working with a system. In other words the Usability engineering community works with a vague and implicit culture bound understanding of usability and its associated methods. While it might seem logical for US military personnel to expect effective and efficient use of systems, it must be doubted that this perception of usability is generally valid across professional or (sub)cultural boundaries.

3.2 A Conceptualisation of Usability in Namibia

Usability Engineering is still in its infancy in Namibia. There are neither usability laboratories, nor usability experts or established UE phases as part of the development processes used. Especially large scale development projects such us governmental and parastatal management systems omit the design for and evaluation of usability all together. Only individual software developers integrate

selected usability tasks, such as user prototype evaluations and questionnaires [17]. Considering the low priority of usability as a quality criteria from the developers side, leading to excessive user training and long term help desk activities, it is of utmost importance to establish locally valid UE standards and guidelines. Firstly, the meaning of "usability" in the Namibian context has to be established. Secondly, valid methods have to be evaluated and determined.

In an attempt to ascertain a local meaning of "Usability" a number of investigative sessions with different Namibian user groups were run. The user groups consisted of a number of three to six participants, with variations in gender, age, profession and ethnic background. The participants were grouped according to the software they are working with i.e. two ministry payroll system, two university management system, one agricultural decision support system. Additionally two non-software specific group sessions were held. All sessions were structured in the same manner: First, participants were asked to brainstorm on associative and related terms/concepts of the word "usability" in general. Second, participants elaborated on general characteristics of a "good working environment". Third, participants selected only the appropriate terms from the two previously produced lists which should apply to their software systems, for it to be considered "usable" [17].

Terms that were named most often were: easy, safe, comfortable, specific, reliable, right pace, goal-oriented, and conducive. Interestingly none of the groups mentioned terms commonly associated with usability such as speed, learnability, memorability, or error rates. However a diversified understanding of satisfiability was expressed such as: beneficial, transparent, stress-free and flexible.

This confirms our hypothesis that usability has a completely different connotation in Namibia. However the currently available data is insufficient to determine whether there is a Namibian concept of usability or a user group specific or even individual only. Further data will be collected for more detailed statistical evaluations.

Furthermore, this investigation shows how differently assumed-to-be-usable systems should be designed and evaluated. Developers can no longer rely on their professional intuition and assumptions but actively and explicitly have to confirm the contextual meaning of the quality criteria with the relevant stakeholders. We have successfully run one of the conceptualisation sessions as part of a participatory design workshop with the client to determine valid evaluation mechanisms. This will support the development of a usable system in the terms of the client.

4 Culture-Centric Development (CCD)

Commonly seen as part of globalisation efforts, UE is mistakenly understood to be responsible for a last phase application make-up for foreign markets. This usually encompasses solely the translation of units and layout features, evaluated by

some – if any – cultural representatives residing in the country of development. In few cases the system is evaluated in the local context with real end-users and standardised, predefined methods. However such a small and late involvement of end-users has been long criticised. New Software Engineering paradigms like user-centred and agile development, interaction and participatory design, established the relevance of user concerns and early involvement. Thus UE activities should commence in or before the first phase of software development processes.

Especially in a cross-cultural development a thorough understanding of the cultural context has to be acquired to guide the development process as well as design decisions [10]. As culture is subject to constant changes within its defining environment, we need to account for this fact by adjusting the applied methods and evaluation tools. While the need for such adjustments is well-accepted in SE, the same has to apply to UE as well. We are therefore proposing the incorporation of arbitrary SE process models in an extended framework, embracing the underlying culture at all times (figure 2).

4.1 A Framework for Culture-Centric Design

Definition of Quality (Usability) Criteria: In the first contact phases with the client the quality criteria of the system should be negotiated. This includes the explicit definition of usability within the context which involves intensive user participation. An example which we have already successfully applied are conceptualisation sessions in the form of workshops as described in 3.2.

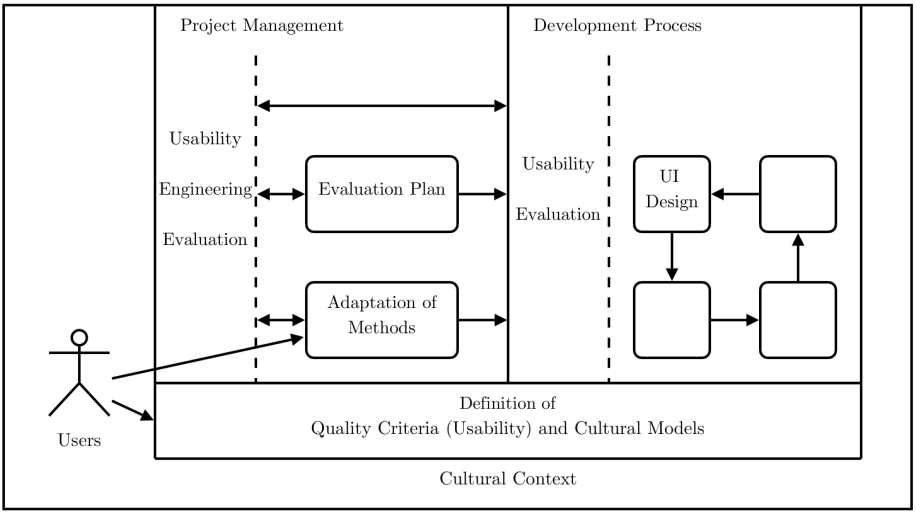


Fig. 2. The CCD framework is connecting Project Management and the Development Process (using an arbitrary development process model) through continuous usability evaluations within the cultural context

4.2 Project Management

Adaptation of Evaluation Methods: In an early design phase, an assertion of explicit quality metrics to judge the system once delivered needs to take place [2]. Based on the identified metrics and evaluation criteria, a local acculturation of methods will have to be done. Again, this acculturation must incorporate user input as well as cultural models.

Evaluation Plan: Based on the identified evaluation methods, an evaluation plan is established, stating implementation details and the assessment criteria to be applied during the evaluation within the development process.

Usability Engineering Evaluation: While software development processes are already subject to project-specific selection and customisation, the additional UE tasks as proposed are designed to be independent of the SE process model as such. Thus, the complete UE process should be evaluated in the scope of a higher-order evaluation taking place within the project's management. Although we deem the suggested CCD framework applicable and highly useful in its current form, it shall be subject to continuous improvement. Results acquired from the UE evaluation will therefore be a valuable source of information for future enhancements and refinements.

4.3 Development Process

While development process models are usually chosen on a per-project basis, we suggest to prefer models embracing change in all its varied forms and allow for high frequency iterations. Agile development, Extreme Programming and prototyping in general seem more applicable in the CCD framework as they give users a deeper insight and thus allow for a closer co-operation throughout the development.

Usability Evaluation: As phases and cycles in a development process are followed as necessary, a continuous evaluation according to the defined usability evaluation methods is taking place, feeding the outcomes back into the process. Furthermore, the evaluation process itself provides input to the UE evaluation controlling the overall applicability and appropriation of the selected methods.

User Interface Design: The design of user interfaces needs to be derived not only from cultural models of how people work and communicate, but also from project-specific guidelines and other locally applicable principles.

4.4 Culture-Centric Development in Western Cultures

The application of the proposed framework is by no means limited to non-western countries. In fact, we suggest its establishment in western settings as well. So far,

usability engineers tend to consider culture as an important factor only if it is not a western one. This leads to the paradox that, although most research in the field of cultural impacts on Software Engineering is done in western countries, developments within these cultures hardly ever incorporate any of the findings of this research. Thus we deem this approach a very valuable one as it would allow to either validate or falsify many of the assumptions used in the majority of Software Engineering projects world-wide. A valid re-definition of usability in developed countries may prove more difficult as most computer literate people already have a heavily influenced concept of usability in mind. Therefore, conceptualisation sessions might not be adequate tools in developed countries.

5 Conclusion

Current internationalisation and localisation efforts are still unsatisfactory in terms of facilitating the design of locally adequate and usable solutions. The lack of empirical studies to inform cultural adaptations of methods and user interface design has to be pursued by the international Usability Engineering community to establish a catalogue of best practices.

Besides, standard usability evaluation encompasses a twofold bias: Initially, through the definition of usability according to western standards, and secondly, through established methods which aim to test an already biased objective. The very foundations and universality of "usability" as it is understood today is doubtful. Conceptualisation sessions held with different Namibian user groups confirmed a deviating perception of the term usability. Thus the concept itself has to be redefined in conjunction with the users to fit the cultural context of the software development and application.

The incorporation of these newly defined UE tasks into existing Software engineering models however leaves us with new challenges, namely the evaluation of the new process. How can we assure that the methods chosen and adapted measure the usability as newly defined and specified within the context, and how can we obtain feedback other than through the long run use of the deployed system?

As Aaron Marcus observes, "*we have barely begun to discover the startling and currently unresearched assumptions about metaphors, mental models, interaction, and appearance. [...] We have an interesting and challenging time ahead of us as we explore the full meaning of cross-cultural user-experience development*" [18].

Acknowledgements

We would like to thank Dr. Sarala Krishnamurthy and Dr. David Cook for their help with the editing of this paper. Special thanks go to Dr. Manfred Meyer for some valuable ideas in the early stages of this paper's preparation.

References

1. Vatrapu, R., Pérez-Quiñones, M.A.: Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. Journal of Usability Studies 1, 156–170 (2006)
2. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction, 3^{rd} edn. Pearson Education, Harlow (2004)
3. Del Galdo, E., Nielsen, J.: International User Interfaces. John Wiley & Sons, New York (1996)
4. Walton, M., Vukovic, V.: Cultures, Literacy, and the Web: Dimensions of Information "Scent". Interactions, pp. 65–71 (2003)
5. Trillo, N.: The Cultural Component of Designing and Evaluating International User Interfaces. In: Proceedings of the 32^{nd} Hawaii International Conference on System Sciences (1999)
6. Hofstede, G.: Cultures and Organizations: Software of the Mind. McGraw-Hill, New York (1997)
7. Marcus, A., Gould, E.: Cultural Dimensions and Global Web User-Interface Design: What? So What? Now What? In: Proceedings of the 6^{th} Conference on Human Factors and the Web, Austin (2000)
8. Ford, G., Gelderblom, H.: The effects of Culture on Performance Achieved through the use of Human Computer Interaction. In: Proceedings of SAICSIT, pp. 218–230 (2003)
9. Fitzgerald, W.: Models for Cross-Cultural Communications for Cross-Cultural Website Design. Technical Report Published as NRC/ERB-1108. NRC-46563, National Research Council Canada (2004)
10. Winschiers, H.: Dialogical System Design across Cultural Boundaries. PhD thesis, Fachbereich Informatik, Universität Hamburg (2001)
11. Winschiers, H., Paterson, B.: Sustainable Software Development. In: Proceedings of SAICSIT 2004, pp. 111–113. ACM Press, New York (2004)
12. Allen, B., Buie, E.: What's in a Word? The Semantics of Usability. Interactions, pp. 17–21 (2000)
13. Shneiderman, B., Plaisant, C.: Designing the User Interface. Strategies for effective Human-Computer Interaction. 4^{th} edn. International Edition. Pearson Education (2005)
14. United States of America, Department of Defense: MIL-STD-1472F. Human Engineering. F^{th} edn. (1999)
15. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human-Computer Interaction. John Wiley and Sons, New York (2002)
16. Badre, A.: Shaping Web Usability: Interaction Design in Context. Addison Wesley, Reading (2002)
17. Stanley, C.: Usability Evaluation of Software Applications in Namibia. Department of Software Engineering, Polytechnic of Namibia, Windhoek. Unpublished (2006)
18. Marcus, A.: Culture: Wanted? Alive or Dead? Journal of Usability Studies 1(2), 62–63 (2006)

Analyzing Non-verbal Cues in Usability Evaluation Tests

Pradeep Yammiyavar[1], Torkil Clemmensen[2], and Jyoti Kumar[1]

[1] Indian Institute of Technology, Guwahati, India
{jyoti.k,pradeep}@iitg.ernet.in
[2] Department of Informatics, CBS, Copenhagen, Denmark
tc.inf@cbs.dk

Abstract. Verbal data is the primary focus for analysis in the prevalent Usability evaluations like in 'Think Aloud Method'. This study involves 18 cross cultural TA tests and it was found that users use gestures profoundly to communicate their mental activities. It was observed that hand gestures are attempts to communicate abstract feelings as well as to quantify, to simplify a complex expression & refer to fuzzy thoughts. 10 further TA tests, with close up cameras for capture of facial expressions yielded gestures of affect states of surprise, satisfaction, confusion, deep thinking, frustration and boredom being experienced by the user. Most importantly, the users were either verbally silent or were using words seemingly incongruent to verbalisation. Observing that there is rich meaning in gestures, this paper argues for gestures as additional data sources in TA analysis.

Keywords: Think Aloud Test, Gesture Analysis.

1 Introduction

Verbal protocol analysis has become an accepted tool for usability evaluation in HCI field. Think aloud (TA) as a method of understanding the cognitive processes [1,2,3] of the user is being used extensively in both academic research and industry applications. Inadequacies of verbal expression and mismatch of verbal capacity and fluency to the speed and complexity of thought processes have also been reported [4,5]. Non verbal cues like gestures, eye movements and tonal variations are observed in the user's attempt to express what is going on in his/her mind during the task fulfillment in the think aloud usability tests [6]. Each gesture or movement can be a valuable key to recognize emotion a person may be feeling at a time. What people say is not always what they mean or are feeling [7]. According to Korchin [8] a gesture can seem as an intentional act of communication. Gestures along with bodily movements, postures, gait, facial expression and non verbal speech patterns can unintentionally yield information. Way back in 1968 Mahal [9] found that personally meaningful gestures reappeared periodically during interviews. Some movements had the same meaning and occurred simultaneously with verbal activity.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 462–471, 2007.
© Springer-Verlag Berlin Heidelberg 2007

In the conducted TA tests, the selected sample users sat before a computer in a standard usability testing setup to fulfill a task predetermined by the tester/ testing team. The subjects were asked to Think Aloud while he/she performed the task/tasks. The users were expected to be or get acquainted with the thinking aloud method as he/ she received instructions and performed a mock test. The tests were done in two stages. Having observed in stage one, the significance of gestures, stage two experiment was conceived to study the phenomena deeper. The interaction of the facilitators and users were recorded in the first stage of the experiment whereas both the interaction and the facial expressions of the users were captured in the second stage done after a gap of 9 months. 18 facilitator-user pairs of which 6 pairs consisted of facilitators from European origin and users from Indian origin were involved in the first stage of the tests whereas 10 facilitator-user pairs from the same culture took part in the TA tests in the second stage. The task in the first stage was to make an invitation card for an Indian wedding whereas the task in the second stage was to select a place of choice for picnic with friends in national tourism websites of three countries, namely, India, China and Denmark. The Recorded videos in the first part were coded for most occurring gestures and the frequency and intent of each gesture as understood by same culture two interpreters was registered. In the second stage images of gestures form the videos were extracted and shown to the acquaintances of the users for their interpretation. Finally a list of identified important gestures has been made and is being reported in this paper.

The use of non verbal cues to understand the cognitive processes of the users more reliably in the usability tests is being suggested in light of the findings from the experiment. Non verbal cues (NVC) were found in user's recorded behavior occurring repeatedly over several think aloud tests. The number of gestures was found to increase when the user was groping for words or had no words to describe a certain mental state like frustration, recall etc.

2 Method

The think aloud usability tests were conducted in two stages, in first stage, interaction behaviors and gestures of the facilitators and the users were observed at a gross level and in the second stage close up facial expressions of users were focused at. The second test was conceived after the findings from the first stage suggested the powerful role gestures were playing in the user's communications during TA.

2.1 Stage One : Think Aloud Tests with Interaction Analysis Setting

In this stage, 18 think aloud usability tests of a task on a software known to the users with seeded errors for the task were conducted with different user - evaluator pairs. The pairs varied in hierarchy, age, gender and country of origin. 6 evaluators; 3 bachelor degree Indian students (age group 19-21 yrs); 2 European academicians and 1 Indian academician (age group 35-50yrs), each having evaluated 3 different users for the given task using think aloud method were video recorded. The task was to design a wedding card invitation for an Indian marriage. The camera was placed so as to record the interaction of the facilitator-user pair.

First, the recorded videos were qualitatively analysed by replaying them several times for identifying relevant verbal and non-verbal behaviours of users and evaluators. Next, the non verbal behaviours were identified and were checked for their frequency. For instance, in table 4.1, subject D1, the user was raising his hand often to show something he intends to say about the images being used in the wedding invitation card making. These often repeated gestures were then registered and images were extracted.

Two independent reviewers (one in late '40s, the other in 20's, both male) were asked for their interpretation of the images. It was difficult to make meaning out of the hand gestures initially by just looking at the images. Hence the movie clips were made and then were analyzed. The mutually agreed interpretations were registered. Some of the images are presented in Table 1.

2.2 Stage Two: Think Aloud Tests with Facial Expression Recording Setting

Ten facilitator-user pairs where all were acquaintances for last three years, age group 20-23 years, nine male-male pairs, one female-female pair, did the Think Aloud test for website exploration of three countries, looking for places to visit with their friends on the national tourism websites of China, India and Denmark. The facilitators were trained during a course in usability on how to conduct TA tests. A scenario for the task was presented and then the task was introduced to the users. The users were given an approximate time of 45 minutes for all the three websites, with a minimum time limit of 10 minute per website and a maximum of 20 minutes. At the end of each test a qualitative interview regarding the interactions was conducted wherein questions regarding the interaction, TA behaviour, the satisfaction level and impact of this method on their task fulfillment were asked and then developed upon to get more insights. A close up camera captured the facial expressions of the users along with the think aloud behavior. The verbal data was recorded along with the screen capture of the activities. The recorded video was coded for the gestures. The most repeating facial expressions were registered. Also strategically most important gestures involving facial expressions were registered. Images of these identified gestures/ facial expressions were extracted from the video.

The extracted images from stage two were shown to a) the 14 acquaintances of the subject, who had known him/her for more than 3 years b) strangers to the subject but from the same age group and culture c) the user himself/herself. The subject was shown the images after 3 days of the conclusion of the tests. Because the facilitator was an acquaintance himself/herself, his/her views were also taken along with separately. The subjects, acquaintances and the strangers all were asked the same question, " What do you think, the person's expression is." All the acquaintances had undergone the tests and hence they knew the context of the images. Three reviewers who had not known the subjects (age group 20-35) were asked for their understanding of the expressions without telling them the context.

3 Results

In stage one, 14 out of 18 videos (4 were discarded for technical quality reasons) have been analyzed for gesture interpretation in form of images and clips. Images and clips

were shown to the independent reviewers from the same culture. Presented in Table 1 are some of the most occurring and strategically most relevant images of the gestures, for example. The Think Aloud behavior along with the gestures is also presented side by side. The mutually agreed interpretation by the reviewers is also presented in the last column consisting of the researchers' inferences from the gestures.

Table 1. Most occurring and strategically relevant images of the gestures

Case	Gesture Image	Behavior	Interpretation of gesture	Reason for the gesture
D1		"…the images on screen are small…"	Hand Gesture used to focus on the size of images and show by gesture that they are this much small.	The amount of smallness and resulting uneasiness has to be expressed..
D2		"…these Westernised Kind of …."	The invitation card has feel of western culture…the hands attempt to communicate what is difficult by words- namely western orientation.	Subconscious attempt to communicate his feel of western, which is very abstract
D3		"…it starts with a paragraph…"	The staticness of the paragraph over images or free flowing text is being intended in the card design	Communicating the abstract feeling of the staticness
D4		"…the bold letters.."	In the card design, he is thinking about the impact of the bold letters and is building his words to communicate the feeling.	Thinking and drawing attention to being surprised with the boldness effect.
D6		"…. With the figure…."	Explanation for the part of the figure that is being referred … the curvy movement of the hand shows the intended feature of the image that makes him like it	To simplify the composites of image and show the relevant element of the image that is in the context
D7		"…in Indian tradition, things like rangoli…"	Explaining to a western facilitator an Indian concept related to cards and attempts to show through hand movement	Introducing a concept (rangoli) which is alien to the facilitator hence putting all means of communication at use to make oneself understood.
D8		"..this much width.."	The card needs to be this much wide	It is difficult to quantify, state in words, the amount
D9		"… the background is…hmm…"	Thinking.	Thinking. Looking for a word

In stage two, 8 of the 10 videos (two were discarded for technical quality reasons) were identified for the image extraction of gestures. The most occurring gesture's images were extracted and shown to the a,b and c groups as mentioned above. From the pool of interpretations, common words, only differing in linguistic labels, were grouped and identified in one of the selected words. These words have been presented in bold letters in Table 2 along with the identified images. Other interpretations, not elicited commonly by most reviewers have been tabulated in normal fonts.

Table 2. Interpretations of the facial expressions of the test users

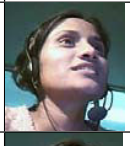
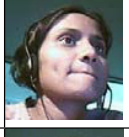
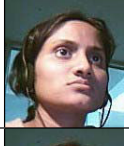
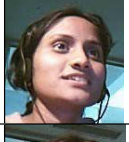
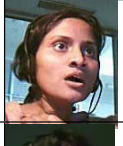
Image	Facial expression	Interpretations			Users TA Behaviour
		A (By acquaintances)	**B** (By Non Acquaintances)	**C** (By User)	
S1.1		**Fascinated**	Observing, Searching, expecting, Curiosity	Fascinated	Silence
S1.2		"Oh, I've got it", **Deciding**	Thinking, Deciding, Waiting, About to express something	Scrutinizing	Silence
S1.3		**Frustration**	Only Looking at something, not thinking much, Something went wrong, Thinking	"not able to get what I want"	Silence
S1.4		**Happy**	Seen some Familiar 'thing' and has some views, Expected result has come,	"Got what I wanted"	" Oh!, it was this" regarding an activity on the
S1.5		**Astonished**	Astonished, Something wrong, Dislike for something	Can't remember	Silence
S1.6		"Oh, What has happened, afraid, surprised	**Astonished**, Something went terribly wrong, surprised	I was bored, was yawning	Silence
S2.1		**Excitement**, Something new, enthusiastic	**Happy**	Excitement	Silence

Table 2. (*Continued*)

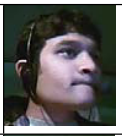
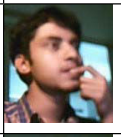
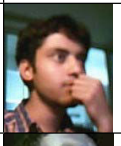
S2.2		**Searching**	**Thinking**, Seriousness	Was Thinking	Silence
S3.1		Thinking, Bored, Load on mind, **Judging**	Observing, **Thinking,** watching seriously	Thinking	Silence
S3.2		**Confused**	**Deciding,** About to say something	Thinking	Silence
S3.3		Have found some way	**Observing,** thinking	Something Interesting	Silence
S3.4		Shy of saying something	**Thinking**	Reading and thinking	Silence
S4.1		Frustrated, **Tense**, Perplexed, in pain due to something, burdened.	**Confused**, Is forced to make decision, Tense, natural	Tired	Inaudible murmur
S4.2		**Observing and Thinking**, Determined after thought	**Worried,** natural	Observing a particular detail on the screen	Silence
S4.3		**Happy**, Got what was being looked for	**Happy**	Looking at a picture on the screen	Silence
S5.1		**Surprised**, 'what is this'	Thinking, Astonished, curious	Was observing	Silence
S5.2		**Inquisitive**, 'what is this'	**Thinking**, 'How can it be'	Observing	Silence

Table 2. (*Continued*)

S5.3		About to speak	Observing, **Thinking**	First time looking at something	Silence
S6.1		Found some known thing	**Happy**	Normal look	Silence
S6.2		What does this mean?	**Observing**	Thinking	Silence
S7.1		**Thinking**	Thinking seriously, trying to make some opinion, observe	Thinking	Silence
S7.2		**Suspicious**	Little Confused, trying to make opinion, observe	Observing	Silence
S7.3		Correlating something	Quite Worried, Confused, observe	Observing	Silence
S7.3		Surprised, **Inquisitive**	Seen a ray of hope	Annoyed	Silence
S8.1		Bored, doubtful	Doesn't agree to what he is watching, observe	Observing	Silence
S8.2		Concentrating	Watching, normal	Thinking	Silence

4 Discussion and Conclusion

We have five observations from the analysis of gestures and facial expressions of test users during think aloud usability tests:

Table 3. Common facial gestures among eight test users in usability tests

Subject	Most Recognizable facial Gestures Elicited
S1	Surprised, Deep thinking and Bored
S2	Happy and thinking
S3	Confused
S4	Happy, Worried and Puzzled
S5	Frustrated and Happy
S6	Surprised and Inquisitive
S7	Thinking and Happy
S8	Concentrating

1. Subjects varied in kind of gestures made though a few gestures were common to all, namely, happiness, boredom, frustration and deep thinking (Table 3). Also the number of gestures made were different in different users. For instance the subject S1 and S7 had shown more gestures as compared to S3 and S8. Also the think aloud behaviour was more in duration and came naturally to some of the subjects without they being aware of it, like S1 and S9 while it was very little in the case of S3 and S7. The individual differences in the TA behaviour and the gestures have been found to compensate for the verbalisations in the usability tests in the case of S7 where more gestures and less TA behaviour was observed and accentuate in case of S1 where both were in greater magnitude comparatively making the subject more easy to analyse for the satisfaction in interaction with the website. In cases of other subjects (S2-S6 and S8-S9) the gestures and verbal data enrich each other, sometimes complementing (when both gesture and think aloud behaviour is there) and on other occasions supplementing (when the subject is silent, but there is/are gesture/s or vice- versa). In case of subject S3 the gestures become crucial for there is less of TA and less of gestures. For thinking aloud doesn't come naturally to the user and by less gestures here we only mean less no. of gestures, for the person is always in some position and the *face can still be the index of mind*, the only issue is of the agreeing upon the meaning of the expression.
2. We observed that 5 out of 10 subjects reported that the TA behaviour interfered with that part of the task which involved thinking and deciding from several choices, for example, which hotel to book, depending on the aesthetics, cost, location etc. Our inference in this case is that the cognitively loaded tasks demanding more of mental resources and hence making the Think aloud behaviour an interference with the task itself. This has been an oft reported phenomena related to think Aloud and these findings only substantiate it. Therefore it is posited that gestures are brought into play to reduce the cognitive load.
3. Subjects (6 out of 10) reported the exhausting effect of think aloud. They reported that the TA activity made them tired and 2 of them felt even hungry at the end of the tests. Our inference here is that the extra effort spent in thinking aloud, specially when a facilitator keeps on nagging to keep thinking aloud may also have physiological effects on the body of the subject, in terms of stress, strain and fatigue. And it is highly probable specially in cultures like India that these

physiological loads may be expressed more through gestures rather than verbally, similar inferences have been reported by other researchers [7].

4. It was sufficient in the facial expression inference case to just show the image (except for the case of yawning) while the hand gesture interpretation required longer movie clips of the gestures in order to make judgement about them. It seems that the hand gestures involved spatial and time domain activity and were used to accompany the verbalisation while the facial expressions were more tacit, and affect related and were *' there on the face',* hence only images gave good results and satisfaction was reported by the interpreters in case of facial expressions.
5. The interpretations of the facial expressions shown in the images were found to give cue to almost same understanding of the gestures by the user (except for the case of yawning image, S1.6) and their acquaintances, whereas the strangers' interpretations had deviations from reports of the user. The user's and facilitator's reporting, upon showing the image was more in terms of what the person was doing at that instance, than what the expression meant. This is a positive observation towards developing cultural gesture recognition protocol, for the acquaintances are able to predict the expression in the image. The failure on the part of strangers on the other hand challenges such a possibility, as they too are from the same culture but are not able to identify the expression in the image.

To conclude, from the above series of observations and inferences it is clear that gestures carry substantial amount of information which can be tracked to enrich/ complement or supplement the verbal data obtained during the usability tests aimed at knowing the cognitive processes of the users mind. Researchers like Albert Mehrabian [11] have held that transmission of message is effective only when all the three aspects of communication – the verbal (words - 7% impact), the vocal (intonation, pitch ,volume - 38% impact) and the visual (gestures ,postures –55% impact) are in tandem with one another. In light of the above inferences it is posited that some of the nonverbal behaviors expressed through gestures can act as clarifiers of the communication that is happening in a standard think aloud protocol situation.

5 Future Work

The gestures can further be analyzed for cross cultural interpretation of static images of gestures and movie clips of the gesture elicitation. People from other cultures can be subjected to these images and asked what they understand from it. Table 1 and Table 2 can further be detailed out with time duration and strategic locations of the gestures in the usability tests, including the role of silence. Further work can be done towards developing a culture based 'gesture lexicon' which can be used by cross cultural usability testing professionals using TA method.

Acknowledgements. This study was co-funded by the Danish Council for Independent Research (DCIR) through its support of the Cultural Usability project.

References

1. van Someren, M.W., Barnard, Y.F., Sandberg, J.A.C.: The Think aloud method - A practical guide to modelling cognitive processes, p. 26. Academic Press, London (1994)
2. Nielsen, J., Clemmensen, T., Yssing, C.: Getting access to what goes on in people's heads? - Reflections on the think-aloud technique, Nordi CHI (2002)
3. Ericsson, K., Simon, H.: Protocol Analysis – Verbal Reports as Data. MIT, Cambridge (1993)
4. Boren, M.T., Ramey, J.: Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication 43(3), 261–278 (2000)
5. Branch, J.: The Trouble With Think Aloud: Generating Data Using Concurrent Verbal Protocols. In: Proceedings of CAIS 2000: Dimensions of a global information science (2000)
6. Yammiyavar, P., Goel, K.M.: Emphasis on non-verbal cues for interpreting cognitive processes in protocol analysis. In: Proceedings Indo Dan Research Symposium on HCI, Guwahati (2005)
7. Pease, A., Pease, B.: The Definitive Book of Body Language, Pease International Pty Ltd, Australia (2005)
8. Korchin, S.J.: Modern Clinical Psychology; Basic Books, New York (1999)
9. Mahl, G.F.: Gestures and Body movements in interviews. In: Shlien, J.M. (ed.) Research in Psychotherapy, vol. 3 , American Psychology Association, Washinton, DC (1968)
10. Mehrabian, A.: Silent messages: Implicit communication of emotions and attitudes, 2nd edn. Wadsworth, Belmont, California (1981)

Online Analysis of Hierarchical Events in Meetings

Xiang Zhang[1,2], Guang-You Xu[1], Xiao-Ling Xiao[3], and Lin-Mi Tao[1]

[1] Department of Computer Science, Tsinghua University, 100084 Beijing, China
[2] Yangtze University, 434023 Jinzhou, Hubei ,China
[3] Department of Computer Science and Technology, Wuhan University of Technology, 430063 Wuhan, China
{xiang-zhang,xgy-dcs,linmi}@tsinghua.edu.cn, xljrzx@163.com

Abstract. Automatic online analysis of meetings is very important from three points of view: serving as an important archive of a meeting, understanding human interaction processes, and providing the attentive services based on the meeting situation for participants. Based on this view, this paper presents principle and implementation of online analysis of hierarchical events in meeting scenario. A hierarchical dynamic Bayesian network modeling different levels of events is designed. In this model, the recognition of low-level events is supervised by high-level events Rao-Blackwellized particle filter is proposed for on-line inference for the hierarchical dynamic Bayesian network. Situation events and four sorts of interaction events in meeting scenario are detected and recognized. Experimental results show that our approach can detect and recognize multi-layer semantic events in dynamic environment. Comparing with previous methods of meeting analysis, our approach supports online probabilistic inference for activities at different layers in meeting scenario.

Keywords: Meeting analysis, dynamic Bayesian network, particle filter, event detection and recognition.

1 Introduction

Meetings play an important role in social communication and interaction [1]. Meeting minutes can serve as an important archive of a meeting but they can't provide the attentive services based on the meeting situation for the participants, if the minute is derived off-line after the meeting. Further some important events might be missed. Therefore, it is very important to develop an online meeting archive system capable to analyze the hierarchical events during the meeting, so as to response accordingly. Meeting analysis involving group interaction has attracted attention in fields spanning computer vision, speech processing, human-computer interaction, and information retrieval. Sample applications include structuring, browsing and querying of meeting databases, and facilitation of remote meetings.

Meetings recorded in multi-sensor room consist of multimodal streams of audio and video, captured with multiple cameras and microphone arrays covering participants and workspace areas. The semantic approach is based on representing the meaning of multimodal behavior of a meeting participant using information obtained

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 472–479, 2007.
© Springer-Verlag Berlin Heidelberg 2007

from different sources, as well as on recognition of meeting situation actions using semantic features extracted from participants multimodal behavior. Although there are many researches involved in meeting archiving and individual and group actions analysis in meeting rooms, such as Distributed Meeting System from Microsoft [2], Multimodal Meeting Tracker from CMU [3], AMI System from IDIAP [4], and EU research project M4 [5], still few researches tried to incorporate context awareness into meeting event analytical problems. In CMU's CAMEO system [6], personal states and meeting states are inferred through finite state machine models separately, where only users' standing-sitting states are taken into account. McCowan et al. [7, 8] used Layered Hidden Markov Models for the recognition of individual and group actions in meeting scenarios based on audio-visual information. Hakeem and Shah [9] proposed an ontology and taxonomy framework for the offline classification of meeting videos. Al-Hames and Rigoll [10] employed dynamic Bayesian network for the recognition of group actions in meeting video sequences. Most of the research literatures mentioned here performed offline meeting event analysis in predetermined and constrained context models [11,12]. To date most approaches to automatic meeting analysis have been limited to the analysis of the individual actions of meeting participants. Recent work has started to explore multi-person scenarios, where not only individual but also group actions or interactions become relevant.

In this paper, we will describe our semantic approach in modeling a meeting as a sequence of meeting situation events. We propose a hierarchical dynamic Bayesian network for recognition of hierarchical events in meetings. In this model, the recognition of low-level events is supervised by high-level events (regarded as the context). Another novelty of our work here is that we show how Rao-Blackwellized particle filters can be applied to efficiently online estimate joint posteriors over hierarchical events in meeting.

Based on this approach, our paper is organized as follows. Section 2 introduces the meeting room configuration. Section 3 proposes hierarchical events model which we used as hierarchical events structure in the meeting scenario. Section 4 describes the online inference method for hierarchical events. Experimental results are presented and discussed in Section 5, and Section 6 concludes this paper.

2 Meeting Recoding Configuration

The meeting room configuration for events analysis and recordings is illustrated in Fig. 1. Multiple sensors are installed in the meeting room so as to acquire the overall information about the environment in real time. Three fixed cameras are set to extract visual information from three distinct perspectives, where two cameras each acquired a font-on view of two participants including the table region, and the third camera behind the table observes various activities of the participant in font of the meeting room. Three Intel-provided microphone arrays detect the direction of the sound resource and speech activities in real time. The seating positions of four participants were allocated randomly.

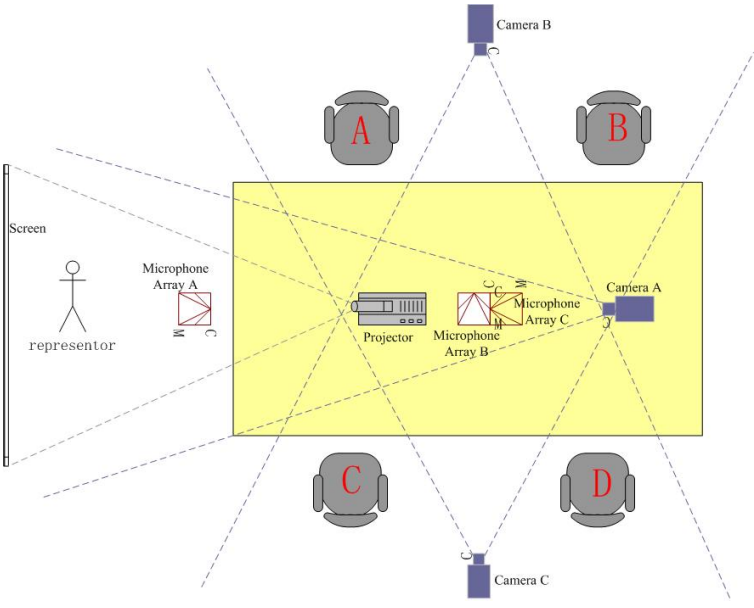


Fig. 1. Meeting recording configuration

3 Hierarchical Events Model

A common drawback in all previously proposed approaches is that they feed the sensor data or features into static classifiers, or a bank of temporally independent HMMs. Further, most of the previously proposed algorithms do not make a distinction between 'complex' and 'simple' activities [13]. In practice, it might be advantageous to decompose complex activities into simpler activities that might be easier to learn. Our approach divides the meeting actions into three hierarchical components: a set of multimodal group actions (situation events), a set of the interaction actions (interaction events) and a set of individual actions (entity events).

3.1 Definition of Events

A group of four situation events are defined based on multi-modal turn-taking patterns in the meeting scenario. The list is defined in Table 1. These situation events are multimodal, non-overlapping and exhaustive, and commonly found in meetings. There are all natural actions in which participants play and exchange similar, opposite, or complementary roles. For examples, during a monologue, one person speaks to the group, while the other participants listen and direct their gaze towards the speaker or to another one. During a discussion, multiple participants take relatively short turns at speaking, and more movement could be expected.

Four sorts of interaction events during four situation actions are respectively defined, such as motion interaction activities, multi-person speaking interaction activities, appearance of person activities in five room areas, and using the projector screen. Some individual actions consist of sitting, standing, speaking, et al.

Table 1. Description of situation events

Situation events	Description
Monologue	one participant speaks continuously without interruption for a long time
Presentation	one participant at font of room makes a presentation using the projector screen
Discussion	all participants engage in a discussion
Break	each participant is free

3.2 Hierarchical Dynamic Bayesian Network

The dynamic Bayesian network(DBN) allows the construction and development of a variety of models, starting from a simple HMM and extending to more sophisticated models, with richer hidden state [14]. Among the many advantages provided by the adoption of a DBN formalism, one benefit is the unequalled flexibility in the model internal state factorization. Situation events and interaction events in intelligent meeting scenario are detected and recognized here. Two-level hierarchical events, combining with vision and audio feature cues, are modeled using the hierarchical dynamic Bayesian network, as is illustrated in Fig. 2.

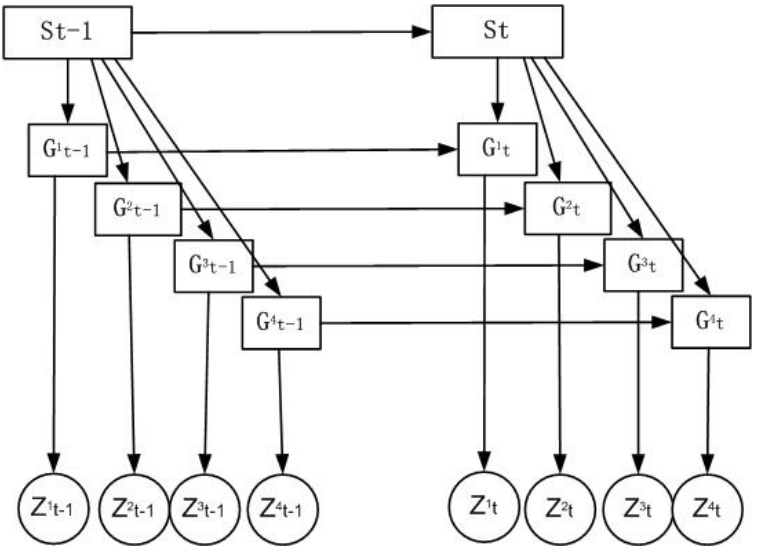


Fig. 2. Dynamic Bayesian network structure for meeting events analysis; square nodes represent discrete hidden variables and circle nodes denote discrete observations

The vision and audio observations of human objects in the five interest areas are treated as observation nodes $Z_t^1 \sim Z_t^4$ of the Dynamic Bayesian Network. Dynamic Bayesian networks estimate four sorts of interaction events $G_t^1 \sim G_t^4$ in the five

interest areas from the given observation, and further infer meeting situation events S_t from four cues of interaction events.

4 Inference

Comparing with previous methods of meeting analysis, our approach supports online probabilistic inference for activities at different layers in meeting scenario. During inference, our system estimates a joint posterior distribution over the complete state space of the hierarchical dynamic Bayesian network. Exact solution to this problem will have exponential complexity in the number of levels of the hierarchical dynamic Bayesian network, and thus is intractable.

We describe how Rao-Blackwellized particle filter (RBPF)[15] can be applied for efficient inference in our hierarchical dynamic Bayesian network. Just like regular particle filters, RBPF represents posterior probability over a state space by temporal sets of weighted samples. RBPF derives their efficiency from a factorization of the state space, where posterior probability over one part of the state space are represented by samples, and posterior probability over the remaining parts are estimated exactly, conditioned on each sample.

According to the structure of the hierarchical dynamic Bayesian network in Fig. 2, the joint posterior distribution $p(S_{1:t}, G^1_{1:t}, G^2_{1:t}, G^3_{1:t}, G^4_{1:t} | Z_{1:t})$ over the complete state space of the hierarchical dynamic Bayesian network for a sequence of T temporal slices can be decomposed as following:

$$\begin{aligned} & p(S_{1:t}, G^1_{1:t}, G^2_{1:t}, G^3_{1:t}, G^4_{1:t} | Z_{1:t}) \\ &= p(G^1_{1:t}, G^2_{1:t}, G^3_{1:t}, G^4_{1:t} | Z_{1:t}, S_{1:t}) p(S_{1:t} | Z_{1:t}) \\ &= p(G^1_{1:t} | Z_{1:t}, S_{1:t}) p(G^2_{1:t} | Z_{1:t}, S_{1:t}) p(G^3_{1:t} | Z_{1:t}, S_{1:t}) p(G^4_{1:t} | Z_{1:t}, S_{1:t}) p(S_{1:t} | Z_{1:t}) \\ &= p(G^1_{1:t} | Z^1_{1:t}, S_{1:t}) p(G^2_{1:t} | Z^2_{1:t}, S_{1:t}) p(G^3_{1:t} | Z^3_{1:t}, S_{1:t}) p(G^4_{1:t} | Z^4_{1:t}, S_{1:t}) p(S_{1:t} | Z_{1:t}) \end{aligned} \quad (1)$$

Our RBPF algorithm samples the situation event variable S, and computes exact posteriors $p(G^1_{1:t} | Z^1_{1:t}, S_{1:t})$, $p(G^2_{1:t} | Z^2_{1:t}, S_{1:t})$, $p(G^3_{1:t} | Z^3_{1:t}, S_{1:t})$ and $p(G^4_{1:t} | Z^4_{1:t}, S_{1:t})$ over the interaction event variables G^1, G^2, G^3, G^4, conditioned on the samples representing situation events.

5 Experiment Results

Observation vectors are formed from a range of audio-visual features that measures the individual events. Audio features, the sound source localization, were extracted from the three microphone arrays. Blobs denoting the human body parts, such as the whole body, head, and face, are extracted from video streams, and represented using boxes. We use box attributes such as width, height, and its center position as vision features to estimate the body poses. We make online recognition of situation events

and interaction events in meeting using the above hierarchical dynamic Bayesian network.

Model parameters are trained using simple counting methods by the annotated meeting sets. For example, the state transition matrix for situation event S and interaction event G^1, G^2, G^3, G^4, or observation matrix is given by:

$$p(x_j|y_i) = \frac{N(x_j, y_i)}{\sum_{s=1}^{M} N(x_s, y_i)} \tag{2}$$

where $N(x_j, y_i)$ is the number of x_j in y_i, M is the number of state x .

Accuracy was determined by counting the number of correctly labeled frames divided by the total number of frames. Three meeting data sets, which contain four situation events, are annotated. By stochastically selecting two meeting data sets as the training data and the rest as the test data. The total recognition accuracy of situation events and four sorts of interaction events is 85.6%. Fig. 3 shows the detection and recognition results for test data for a sequence of 600 temporal slices.

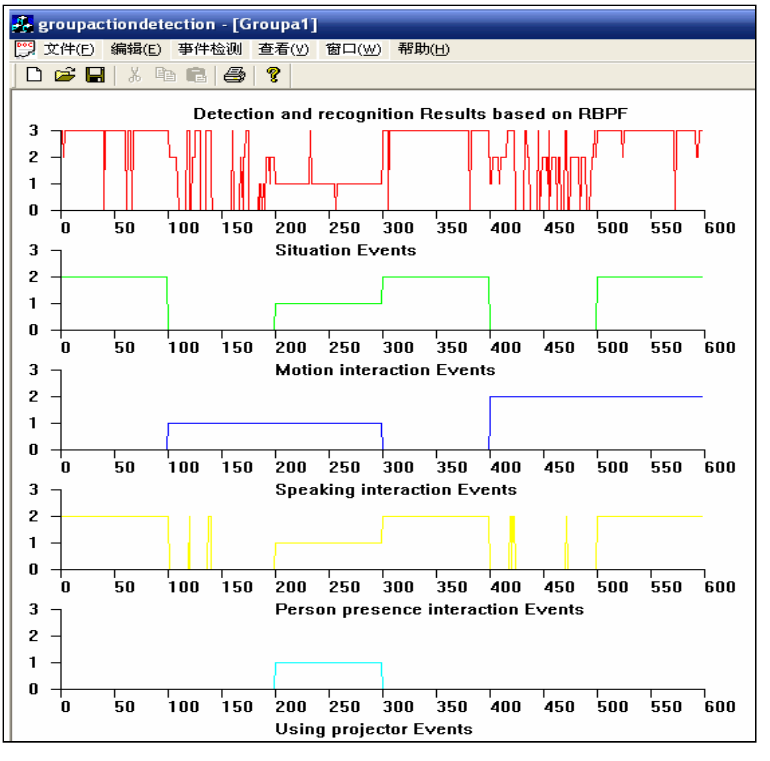


Fig. 3. The detection and recognition results for hierarchical events in the meeting scenario

6 Conclusion

We have described principles and implementation of online analysis of hierarchical events in meeting scenario. We have introduced a hierarchical dynamic Bayesian network that has the ability to model different levels of events and observation features. Rao-Blackwellized particle filter (RBPF) is proposed for on-line inference for the hierarchical dynamic Bayesian network. Some important characteristics of this paper when compared to the previous research works in meeting analysis are: (1) our work can detect not only individual actions but also group or interaction actions, which is very relevant to automatic meeting abstraction; (2) most of the research literatures performed offline meeting event analysis in predetermined and constrained context models, our work can make on-line detection and recognition of human activity in dynamic environment. Experimental results have validated our approach, which show that the RBPF can detect and recognize multi-layer semantic events in dynamic environment.

We are currently exploring several theoretical and engineering challenges with the refinement of hierarchical event analysis in dynamic environment. Context analysis and the relation between context model and event detection are considered for the future extension of our work.

Acknowledgments. This work was funded under Project 60673189 supported by National Science Foundation of China: Event Detection and Understanding in Dynamic Context for Implicit Interaction and Project 2005038351 supported by the Chinese Postdoctoral Science Foundation: On-line Event Detection Based on Context Aware. In addition, we would like to thank the anonymous reviewers for their insightful comments.

References

1. Reiter, S., Rigoll, G.: Multimodal Meeting Analysis by Segmentation and Classification of Meeting based on a Higher Level Semantic Approach. In: Proc. IEEE ICASSP, Philadelphia, USA (2005)
2. Cutler, R., Rui, Y., Gupta, A.: Distributed Meetings: A Meeting Capture and Broadcasting System. In: Proc. ACM Multimedia (2002)
3. Bett, M., Gross, R., Yu, H.: Multimodal Meeting Tracker. In: Proc. RIAO (2000)
4. Carletta, J., Ashby, S., Bourban, S.: The AMI Meeting Corpus: A Pre-announcement. In: Proc. Workshop on Machine Learning for Multimodal Interaction (MLMI) (2005)
5. http://www.m4project.org
6. Rybski, P.E., De la Torre, F., Patil, R., Vallespi, C., Veloso, M., Browning, B.: CAMEO: Camera Assisted Meeting Event Observer. In: Proc. Int. Conf. on Robotics and Automation (ICRA'04), vol. 2, pp. 1634–1639 (2004)
7. McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., Zhang, D.: Automatic Analysis of Multimodal Group Actions in Meetings. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI'05), vol. 27(3) (2005)
8. Zhang, D., Gatica-Perez, D., Bengio, S., McCowan, I.: Modeling Individual and Group Actions in Meetings with Layered HMMs. IEEE Trans. on Multimedia, vol. 8(3) (2006)

9. Hakeem, A., Shah, M.: Ontology and Taxonomy Collaborated Framework for Meeting Classification. In: Proc. 17th Int. Conf. on Pattern Recognition (ICPR'04), vol. 4, pp. 219–222 (2004)
10. Al-Hames, M., Rigoll, G.: A Multi-Modal Graphical Model for Robust Recognition of Group Actions in Meetings from Disturbed Videos. In: Proc. IEEE Int. Conf. on Image Processing (ICIP'05) (2005)
11. Dielmann, A., Renals, S.: Dynamic Bayesian Networks for Meeting Structuring. In: Proc. IEEE ICASSP, Philadelphia, USA (2005)
12. Oliver, N., Garg, A., Horvita, E.: Layered Representations for Learning and Inferring office Activity from Multiple Sensory Channels. Computer Vision and Image Understanding 96, 163–180 (2004)
13. Bui, H.H., Venkatesh, S., West, G.: Policy Recognition in the Abstract Hidden Markov Model. Journal of Artificial Intelligence Research 17, 451–499 (2002)
14. Jensen, F.: An Introduction to Bayesian Networks. Springer, Heidelberg (1996)
15. Murphy, K., Russell, S.: Rao-blackwellised particle filtering for dynamic Bayesian networks. In: Doucet, A., de Freitas, N., Gordon, N.J. (eds.) Sequential Monte Carlo Methods in Practice, Springer-verlag, Heidelberg (2001)

A Cross Culture Study on Phone Carrying and Physical Personalization

Yanqing Cui[1], Jan Chipchase[2], and Fumiko Ichikawa[2]

[1] Nokia Research Center Helsinki, Ruoholahti, 11-13 Itämerenkatu, Helsinki 00180, Finland
[2] Nokia Design Tokyo, Shimomeguro 1-8-1 Meguro-ku, Tokyo 153-0064, Japan
{yanqing.cui,jan.chipchase,fumiko.ichikawa}@nokia.com

Abstract. The mobile phone has become one of the essential objects that people carry when they leave home. By conducting a series of street interviews in 11 cities on 4 continents, we attempted to identify the main carrying options in different cultures and how these options affected user experience in interacting with the phone. We also identified several cultural differences ranging from the prevalence of cases, straps, and other physical phone modification to other ways to personalize and protect the appearance of the phone. Phone straps and decorative stickers were more prevalent in cities such as Tokyo, Seoul and Beijing but seldom witnessed in other cultures. Based on findings from this research, we identified a number of factors that affected carrying position and style, which can be summarized as ease of access vs. the need to maintain security. Non-instrumental attributes include: identity, sociability, and aesthetics. Some practical implications on interaction and industrial design are also discussed.

Keywords: Mobile Phone, Mobile Essentials, Culture, Personalization, Carrying, User Experience.

1 Introduction

The mobile phone is the most ubiquitous Information and Communication Tool (ICT) in modern society and widely considered to be one of the three essential objects that city-dwellers carry with them when they leave home, other others being keys and money. For many people the mobile phone is the first thing that they interact with in the morning, and one of the last objects they use before going to sleep at night. The mobile is typically used in pretty much every context in between. (Chipchase, J. et al, 2005). The baseline functionality that cements the role of the mobile phone in everyday life is its ability to enable personal, convenient, synchronous and asynchronous communication. This assumes however that the user is able to notice incoming communication, but to what extent is this true?

This paper outlines a research study aiming to understand the extent to which incoming communication was noticed by mobile phone users. After the initial study in Helsinki the research goals were extended to include other user experience aspects, and the study was re-run in 10 other cities including Tokyo, New York, Kampala,

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 483–492, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Delhi and Tehran. In addition to answering the initial research question, the results of the studies are also being used to build an understanding of cultural differences in the way users carry and customize their mobile phones, and provide clues that can support the design of wearables and de-converged mobile phones.

2 Previous Studies

The research team has been involved in a number of studies exploring phone carrying behaviours. The first study in 2003 centered on what, why and how people take items with them when they leave home (Chipchase, J. et al, 2005). The qualitative study adopted methods such as shadowing, in-depth interviews and ad-hoc street interviews. Participants for these studies were recruited in Tokyo, San Francisco, Berlin and Shanghai. The study found that the three core items people always carried, regardless of their culture or gender were keys, money and the mobile phone. These items are subsequently referred to as Mobile Essentials (MEs). The study also introduced the concepts of the Center of Gravity to describe where these objects are kept in the home, the Point of Reflection to support remembering MEs before walking out the door and the Range of Distribution – to describe the extent to which objects are allowed to stray from the body, reach and lines of sight when not in use.

The research ignited interest in understanding the nuances of ME carrying behaviors in a variety of cultural settings. Ichikawa and others published initial findings (Ichikawa, F. et al, 2005) noting that most male participants carried their phones in their front right trouser pocket whilst female participants mostly used shoulder or hand bags. The differences between the cities that were studied were not significant.

The authors have observed strong differences in mobile phone use across cultures. For example, users from Japan, Korea and Chinese urban centers often customize their phones appearance using stickers, straps, the extreme cases being part of the Japanese Deco-Den trend (derived from “decoration” and “denwa”, Japanese for “phone”) (Chipchase, J. et al., 2006). In sharp contrast, consumers in the USA and Europe do not personalize their mobile, but maintain it in the same state as purchased. One exception is the purchase of new phone covers, which are still more of a sheath than a customization.

As a part of mobile phone usage practices, the psychological underpinnings of mobile phone personalization are examined by using grounded theories or existing frameworks (Blom, J. et al, 2003; Oulasvirta, A. et al, in press). The motivations of ICT personalization are well associated with basic human needs of autonomy, competence, and relatedness. “<Appearance personalization> is intended to have an effect on other people rather than the user herself”. Physical appearance personalization may serve the functions of emotional expression, ego- involvement, identity expression, and territory marking.

3 Design Research

The research methodology adopted in this study was detailed in a previous paper (Ichikawa, F. et al, 2005). A team of researchers was deployed to conduct street

interview in pairs, one as the interviewer, and the other as photographer. In countries where the research team did not speak the local language, local students were hired and trained to conduct the study. In most of the 11 cities where this study was carried out the research team were already conducting in-depth qualitative data collection – ranging from interviews, shadowing and observations, typically with a small number (<20) participants (Blom, J. et al, 2005). These 11 street surveys provided the research team with an opportunity to meet a wider variety of locals – typically 100+ per city, and get a sense of local tastes and preferences.

The questionnaires used in the study included more then 16 questions, designed to fit on a single A4 sheet of paper. The questionnaire noted the location where the phone was carried, and later where keys and money were carried and the extent to which each item was personalized. Additional questions to probe why were asked. Interviews were conducted in relaxed public settings such as parks and non-busy streets. The team avoided data collection in extreme weather conditions that would bias the type of clothing worn.

As of January 2007, the study has conducted street interviews with 1549 participants from eleven cities in nine countries on four continents. Data collection started in 2003 with Helsinki, New York (NYC) followed by Milan in 2004. Beijing, Jilin, Hyderabad, Tokyo, Los Angeles (LA) and Seoul were done in 2005. Delhi, Kampala, and Tehran were done in 2006. The research team collected at least 50 male and 50 female participants in each city with additional data collected dependent on the availability of local resources. Gender and age of the participants was balanced for each city. Research is ongoing.

4 Phone Carrying Behavior

For research purposes, we defined the phone carrying location as the place the participant currently held the phone, unless the place was identified as being transitional such as their hand. In these occasions, the participants were asked for the usual places they would carry the phone (only on rare occasions the phone was primarily carried in the hand).

4.1 Carrying Options

Generally women used bags and men used trousers pockets as the primary way to carry their phone. The findings confirmed the early conclusion from European cities (Ichikawa, F. et al, 2005). A diagram on general phone carrying locations is shown in fig. 1. The data on individual cities is present in tab. 1.

More carrying options were identified when the study spread from a unified western society into locations where participants came from more diverse cultural backgrounds. For example, the hand was identified as the main carrying option for approximately 6% of all the studied participants, while the neck was the main location for 1%. The hand was identified as the primary phone carrying location in Delhi, Seoul, Jilin and LA. By carrying phone in their hand, people tend to interact with their mobile phone more often.

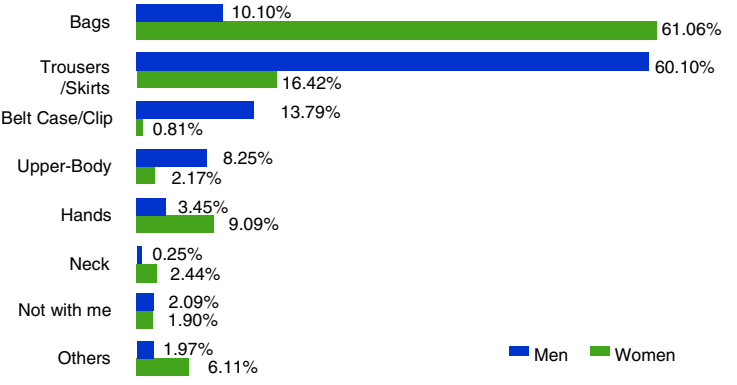


Fig. 1. Location in all the studied eleven cities

Carrying their phones in their bags was a relatively new habit for the women in some studied cities. Approximately 80% of Western (Helsinki, NYC, and Milan), women carried their mobiles in their hand bags but only. 50% or less of their counterparts in less developed cities such as New Delhi and Jilin followed the same practice. In the cities with less prominent culture of using bags e.g. Delhi, trousers pockets were the common carrying location.

Phone carrying is contextually dependent. The study in LA was conducted between Santa Monica and Venice Beach where the contextual factors differed from other cities – for example clothing was more orientated to leisure and beach activities. These contextual differences were reflected in the high difference in carrying options.

Table 1. Gender difference in phone carrying options

Gender		Helsinki	NYC	Milan	LA	Beijing	Tokyo	Tehran	Seoul	Kampala	Delhi	Jilin	Sub Tota
Female	Bags	85.33%	83.08%	79.63%	37.50%	67.33%	66.67%	65.85%	61.22%	52.63%	41.03%	39.80%	61.06%
	Trousers/skirts	1.33%	15.38%	11.11%	16.07%	23.76%	15.87%	7.32%	8.16%	8.77%	30.77%	25.51%	16.42%
	Upper-body	2.67%	-	1.85%	-	0.99%	6.35%	-	2.04%	3.51%	-	5.10%	2.17%
	Hands	-	-	-	16.07%	5.94%	1.59%	7.32%	24.49%	-	26.92%	15.31%	9.09%
	Neck	-	-	-	1.79%	1.98%	1.59%	4.88%	2.04%	-	-	11.22%	2.44%
	Belt case/clip	1.33%	1.54%	-	5.36%	-	-	2.44%	-	-	-	-	0.81%
	Not with me	9.33%	-	7.41%	5.36%	-	-	-	-	-	-	-	1.90%
	Others	-	-	-	17.86%	-	7.94%	12.20%	2.04%	35.09%	1.28%	3.06%	6.11%
	Base	*75*	*65*	*54*	*56*	*101*	*63*	*41*	*49*	*57*	*78*	*98*	*737*
Male	Bags	17.86%	12.94%	14.29%	13.64%	13.51%	24.59%	1.85%	7.69%	1.72%	1.23%	1.92%	10.10%
	Trousers/skirts	42.86%	67.06%	62.50%	54.55%	58.56%	62.30%	66.67%	75.00%	74.14%	74.07%	41.35%	60.10%
	Upper-body	13.10%	3.53%	10.71%	1.52%	6.31%	9.84%	11.11%	11.54%	8.62%	12.35%	5.77%	8.25%
	Hands	-	-	-	9.09%	1.80%	-	-	-	-	7.41%	13.46%	3.45%
	Neck	-	-	-	-	-	3.28%	-	-	-	-	-	0.25%
	Belt case/clip	14.29%	16.47%	3.57%	10.61%	18.92%	-	12.96%	1.92%	8.62%	4.94%	37.50%	13.79%
	Not with me	11.90%	-	7.14%	3.03%	0.90%	-	-	-	-	-	-	2.09%
	Others	-	-	1.79%	7.58%	-	-	7.41%	3.85%	6.90%	-	-	1.97%
	Base	*84*	*85*	*56*	*66*	*111*	*61*	*54*	*52*	*58*	*81*	*104*	*812*

4.2 Incoming Notification

The carrying option had an impact on a person noticing incoming notifications, such as calls or messages. When carrying a phone in trousers pockets, approximately 70% of the participants claimed they always noticed the incoming messages or phone call.

This rate was 50% for the participants keeping their phone in their bags. The difference was also reflected between genders since the bag was the primary carrying option for women, and trousers pockets the primary option for men. Approximately 60% of women claimed that they always noticed their incoming communications. The percentage was 71% for men.

Table 2. Incoming notification under different carrying options

Gender	Notice	Bags	Trousers/skirt	Belt case/clip	Hands	Upper-body	Not with me	Neck	Others	Grand Total
Female	No	26.54%	8.47%	-	15.63%	-	35.71%	-	8.33%	20.26%
	Sometimes	24.88%	15.25%	-	10.94%	37.50%	7.14%	6.25%	5.56%	20.26%
	Yes	48.58%	76.27%	100.00%	73.44%	62.50%	57.14%	93.75%	86.11%	59.48%
	Base	*422*	*118*	*5*	*64*	*16*	*14*	*16*	*36*	*691*
Male	No	24.69%	16.59%	8.57%	14.29%	14.75%	25.00%	50.00%	18.18%	16.40%
	Sometimes	23.46%	12.17%	6.67%	3.57%	9.84%	6.25%	-	9.09%	11.90%
	Yes	51.85%	71.24%	84.76%	82.14%	75.41%	68.75%	50.00%	72.73%	71.69%
	Base	*81*	*452*	*105*	*28*	*61*	*16*	*2*	*11*	*756*

4.3 Carrying Decisions

The decision of where and how to carry a phone was made based on a number of factors. These factors can be categorized into 3 categories. "Instrumental" concerns were those factors that were more practical. "Non-Instrumental" concerns were those factors that were more based on preference or opinion. "Contextual Restriction" included those factors that restricted the number of options available to a user based on their current situation.

Table 3. Factors that influenced phone carrying

	Bags	Trousers/skirts	Belt case/clip	Upper-body	Hands	Others	Grand Total
Instrumental	**56.45%**	**69.77%**	**80.17%**	**78.21%**	**82.95%**	**75.00%**	**67.34%**
Uncategorized esasiness in carrying	*20.81%*	*20.56%*	*38.79%*	*19.23%*	*15.91%*	*16.67%*	*21.57%*
Easiness in fetching the phone	*9.25%*	*26.19%*	*19.83%*	*29.49%*	*35.23%*	*19.44%*	*19.97%*
Noticing the incoming call or msg	*4.62%*	*6.68%*	*2.59%*	*11.54%*	*23.86%*	*9.72%*	*7.07%*
Security and prevention for phone	*19.27%*	*15.11%*	*15.52%*	*14.10%*	*6.82%*	*27.78%*	*16.71%*
Health concerns	*2.50%*	*1.23%*	*3.45%*	*3.85%*	*1.14%*	*1.39%*	*2.01%*
Non instrumental	**3.28%**	**3.51%**	**4.31%**	**2.56%**	**1.14%**	**1.39%**	**3.19%**
Fashion or stylish	*2.50%*	*2.46%*	*4.31%*	*1.28%*	*1.14%*	*0.00%*	*2.36%*
Being discreet	*0.77%*	*1.05%*	*0.00%*	*1.28%*	*0.00%*	*1.39%*	*0.83%*
Contextual restrictions	***30.06%***	***10.90%***	***6.03%***	***11.54%***	***10.23%***	***9.72%***	***17.34%***
Best or no other place	*18.69%*	*8.08%*	*1.72%*	*5.13%*	*5.68%*	*6.94%*	*11.03%*
Phone size fit or not for the option	*8.09%*	*1.76%*	*1.72%*	*2.56%*	*2.27%*	*1.39%*	*4.09%*
Not disturbing ongoing activities	*3.28%*	*1.05%*	*2.59%*	*3.85%*	*2.27%*	*1.39%*	*2.22%*
Others	**10.21%**	**15.82%**	**9.48%**	**7.69%**	**5.68%**	**13.89%**	**12.14%**
Total number of commetns	519	569	116	78	88	72	1442

Factors that were seen as Instrumental concerns included: how easy the phone was to carry, or "Ease in Carrying"; how easy it would be to access to phone to receive incoming notification; how easy it was to answer or retrieve the phone; protecting the phone from dropping, losing, scratching or having it stolen, or "Security and Prevention". Factors seen as Non-Instrumental include: local trends or personal style, or "Fashion and Stylish", disliking the presence of a mobile phone, or "Being Discreet". These factors affected each user, culture or location differently. Factors seen as Contextual Restrictions included the following: no other options, big phone size, and not interfering with an ongoing activity. These factors can change with time

for each used based on what they are doing at the time, and so it would change a user's normal or default behavior.

From the chart in Tab 3, we can see a few trends. "Contextual Restriction" played a significant role for users who preferred carrying the phone in their bag. These users relied on the bag to cluster and carry mobile items. "Easiness in Fetching the Phone" was more often a reason for the participants who chose to carry the phone in their trousers, skirts, upper body clothes, and hand as the main option. "Health Risk" was also listed as the primary reason by 2% of participants. These users usually tried to keep their phone distant from their body.

5 Appearance Personalization

Based on the pilot study, we observed that users were likely to personalize their phone's physical appearance using three mechanisms: covers, straps, and stickers. Cover is any type of bag that used to enclose the phone (fig 1, 1-3). Strap is any kind of add-on items with a string that is placed on the strap hole of the phone (fig 1, 4-7). Sticker is a piece of paper or other item that is pasted onto the phone (fig 1, 7-8).

5.1 Personalization Practice

The practice of using phone covers and straps was studied in 8 of the 11 cities, and the practice of stickers on phones was studied in 5 cities. All studied cities witnessed the usage of cases and straps. Sticker usage were commonly found in all Asian cities but for example was barely present in the LA study. Personalization was generally higher for women than for men. The exception to this is that men are more likely to use straps in Kampala, Tehran, and Seoul.



Fig. 2. Covers, strap, and sticker as physical personalization

Covers were more common used in the regions known for their dusty environment in part caused by unpaved roads. 32% of participants in Kampala used phone cases followed by 11% in Jilin and 9% in Delhi. We hypothesize that cover usage is higher in rural environments. Cover usage was surprisingly common in Seoul perhaps explained by a high societal awareness of bacteria and general hygiene – for example carrier shops often include cleaning stations where phones can be scrubbed, airbrushed and irradiated. All eastern Asian cities witnessed the high popularity of

phone strap and sticker usage with approximately 70% of users in Seoul and Tokyo used straps compared to less than 10% for LA and Kampala.

Table 4. Covers, strap, and stickers in different cities

		LA	Kampala	Delhi	Tehran	Jilin	Beijing	Seoul	Tokyo	Subtotal
Cover	Female	10.71%	33.33%	8.97%	19.51%	20.41%	12.87%	22.45%	4.76%	16.02%
	Male	6.06%	31.03%	9.88%	11.11%	2.88%	2.70%	13.46%	1.64%	8.52%
	All	8.20%	32.17%	9.43%	14.74%	11.39%	7.55%	17.82%	3.23%	12.12%
Strap	Female	16.07%	3.51%	11.54%	31.71%	61.22%	60.40%	69.39%	77.78%	43.65%
	Male	3.03%	10.34%	9.88%	33.33%	33.65%	37.84%	73.08%	57.38%	31.35%
	All	9.02%	6.96%	10.69%	32.63%	47.03%	48.58%	71.29%	67.74%	37.26%
Sticker	Female	-%		5.13%	14.63%			12.24%	38.10%	13.94%
	Male	1.52%		1.23%	1.85%			-%	18.03%	4.46%
	All	0.83%		3.14%	7.37%			5.94%	28.23%	8.99%

5.2 Reason Analysis

The reasons for personalization can be categorized as either being more practical, "Instrumental", or more subjective, "Non-instrumental". Some examples for Instrumental reasons are "Usability" related meaning the ease of performing certain tasks, such as pulling the phone from your bag. "Security" of the device is another Instrumental reason. Examples of Non-instrumental reasons include "Aesthetics" and look of the phone, "Identity" and "Sociability", how a user to promote themselves or their group affiliation.

Table 5. Primary reasons in using phone physical personalization

	Cover *(F: 38/M: 18)*	**Strap** *(F: 152/M:123)*	**Sticker** (F: 32/M: 10)
Instrumental	**79%/73%**	**46%/55%**	**6%/30%**
Usability- Easiness in fetching, carrying, cleaning	*3%/0%*	*36%/40%*	*3%/0%*
Security- Protection from scratch, dust, sweat	*71%/67%*	*/*	*0%/10%*
Security- safety from loss, drop, theft, and robbery	*5%/6%*	*10%/15%*	*/*
Security- Protection of privacy in public place	*/*	*/*	*3%/20%*
Non instrumental	**19%/17%**	**61%/46%**	**96%/50%**
Sociability- Received as gift from others	*3%/0%*	*24%/19%*	*3%/0%*
Relatedness- representing thing, moment, or person	*5%/11%*	*16%/5%*	*59%/40%*
Indentity- Changing things into my style	*/*	*5%/3%*	*25%/21%*
Aesthetics	*11%/6%*	*21%/19%*	*12%/-%*
Other reasons	**5%/6%**	**1%/11%**	**6%/20%**

Tab. 5 detailed the reasons for personalization. Covers were more likely to be driven by Instrumental needs, especially for security purpose; stickers were driven by Non-Instrumental purpose. The use of a strap was balanced between Instrumental and Non-Instrumental factors.

6 Discussions

6.1 User Experience Attributes in Carrying

The mobile phone is a portable item and is used in various contexts. As a result, the perspective of how it is carried is essential part of the mobile phone user experience,

which can be studied through carrying options and decision making process. Some aspects are also reflected on how it is personalized.

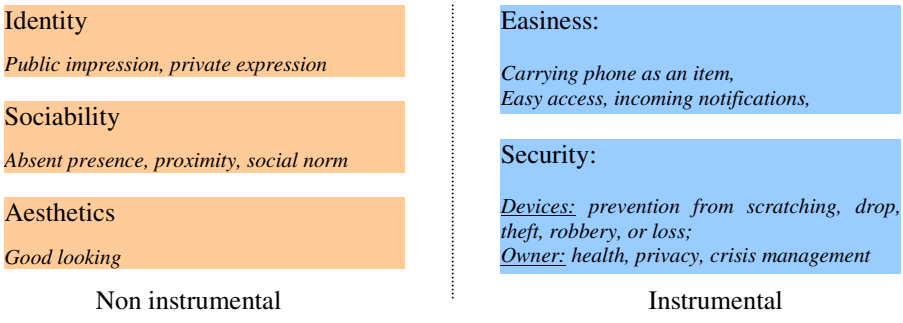


Fig. 3. User experience framework: phone as a carried item

The human practice of carrying a mobile phone is a compromise process between emotional and instrumental purposes. The first instrumental attributes was ease of use, as a carried object and a communication tool. E.g. phone is carried in bag since it is easy to transport, take out although this is compromised by the limited chance of noticing vibration for incoming calls. Next instrumental attribute concerned security, both in terms of the device and its user. E.g. participants used phone cases to protect it from scratch or theft; the phone was placed far away from body since the radiation was perceived as a health risk.

Three non-instrumental attributes were identified in the project. "Identity" addresses the aspects that phone was a way of impression management in public space. "Aesthetics" is a very much related to Identity. "Sociability" refers to how the phone is used for social associations. E.g. phone personalization items were given as a gift. It was common in Seoul for couples to use matching straps.

6.2 Cultural Differences

There were a number of cultural differences in how users personalized, or did not personalize, their phones and how they carried phones. Generally, Asian participants were more likely to physically personalize their phones using straps or stickers than their counterparts in Europe or US. There were also differences in carrying styles - more Asian users carried their phone in hand, for example.

There are a number of different aspects that can explain these cultural differences. The theory candidates range from national culture of dimensions to social context examination. In the theories of cultural dimension, individualism- collectivism can be used in explaining the regional difference well (Hofstede, G. 2004). In a collectivist culture, such as eastern Asia and some part of Africa, people are more likely to create, show, and treasure the association with other people, especially people with strong social-ties such as families. The people also care more about their impression in public. Phone personalization serves as a platform to facilitate their social association and impression management for these cultures.

Other theories such as design culture evolution and social-economic development may also be useful in explaining the regional difference. In economically developing countries, people often place covers on their consumer electronics simply to prolong the life of that product, and to retain its value for possible resale.

6.3 Design Implications

The phone is designed for the primary purpose of synchronous and asynchronous communication. However, in our project, we found these fundamental functions were compromised by the limitation of carrying options. Generally 30% of men and 40% of women do not always notice the incoming calls or messages. The figure was particularly high for bag users i.e. mostly women, at over 50%.

The profile feature on phones can be a useful solution by providing different phone settings for when the phone is in different locations, such as an in-bag profile. The profile features are created for the purpose of (i) avoiding call handling by accident when the user is not aware, (ii) Alternative notification mechanisms to ensure immediate response, (iii) easiness in fetching phone from carrying option, (iv) communication initiator being timely notified about the possible delay.

7 Conclusions

Where a phone is carried is an important part of understanding the total user experience. By conducting a series of street interviews in 11 cities, we tried to identify the main carrying options in different cultures to understand how these options influenced user interacting with the phone. The project confirmed our initial finding that women tend to use bags and men use front (right) trousers pockets as the primary means of carrying their mobile phone. Different carrying options would affect the user's ability to notice incoming calls or messages, with incoming calls frequently being missed when carried in a bag.

The project also identified cultural differences in using phone covers, phone straps, and stickers to personalize the physical appearance of a mobile phone. Phone cover use witnessed in the regions where phone were used in dusty environment, which in turn effects the quality of the user interaction. Phone strap and sticker usage were more often used in Asian cities, especially eastern Asian cities whereas stickers were seldom witnessed in studied in American or African cities. We applied the cultural dimension of individualism and collectivism to explain the regional differences. People from collectivism cultures customize their phone's appearance more often because they are more likely to be used as a platform to create, show, and treasure the association with other people, especially group with strong social-ties such as families.

Based on the findings from phone personalization and carrying behaviors, we identified two types of user experience attributes concerning carrying: Instrumental and Non-Instrumental attributes. Instrumental attributes include: ease of use and security, the Non-Instrumental attributes include: identity, sociability, and aesthetics. The finding of this study can also be used for interaction and industrial design work, and was discussed in the paper.

Acknowledgements. We would like to thank the research team members who conducted the Where's the Phone Studies: R. Grignani, J. Y. Jung, T. Stovicek, Marila, M. Silfverberg, T. Vaittinen, T. Nyyssonen, V. Lantz, and T. Kaaresoja of Nokia; M. Maulini and J. Jalan in Italy; P. Kyungsu and B. Thurston in USA; L. Bitra, S. Sain, P. Sharma, A. Toipathi, S. Swain, S. Singhal in India; S. Lee, M. Chun in South Korea; Y. Zhou, H. Li, T. Zhang in China; Tulusan I. in Uganda. The first author also owes special thanks to T. Stovicek, who did the final draft review and greatly improve the language.

References

1. Blom, J., Chipchase, J., Lehikoinen, J.: Contextual and cultural challenges for user mobility research. Commun. ACM 48(7), 37–41 (2005)
2. Blom, J., Monk, A.: Theory of Personalization of Appearance: Why Users Personalize Their PCs and Mobile Phones. Human-Computer Interaction 18(3), 193–228 (2003)
3. Chipchase, J., Persson, P., Piippo, P., Aarras, M., Yamamoto, T.: Mobile essentials: field study and concepting. In: Proceedings of the 2005 Conference on Designing For User Experience (San Francisco, California, November 03 - 05, 2005), New York, NY, 57 (2005)
4. Chipchase, J., Jung, Y., Heathcote, C., Shimizu, A.: Super Customisation: Deco Den Mobile Phone Customisation in Japan. Nokia Internal Technical Report (2006)
5. Hofstede, G.: Cultures and Organizations: Software of the Mind. New York: McGraw-Hill, USA (2004)
6. Ichikawa, F., Chipchase, J., Grignani, R.: Where's the phone? A study of Mobile Phone Location in Public Spaces. In: Proceedings of the IEE Mobility Conference 2005 (Mobility '05) (Guangzhou, China), 3-2B-2 (2005)
7. Oulasvirta, A., Blom, J.: Motivations in Personalization Behaviour. Interaction with Computuer (In Manuscript) (In Press)
8. Swallow, D., Blythe, M., Wright, P.: Grounding Experience: Relating Theory and Method to Evaluate the User Experience of Smartphones. In: Proceedings of the Annual Conference of the European Association of Cognitive Ergonomics (EACE '05), pp. 91–98 (2005)

Performance Modeling Using Anthropometry for Minority Population

V. Gnaneswaran and R. R. Bishu*

Department of Industrial and Management Systems Engineering
University of Nebraska, Lincoln
rbishu@unl.edu

Abstract. The purpose of this study is to develop predictive models for grip strength, dexterity and manipulability, for four minority populations using anthropometry. A total of sixty subjects representing Hispanics, African Americans, Asian Indians and Vietnamese participated in this study. Subjects performed the three tasks for the following five hand conditions: bare hand, cotton gloves, Kevlar gloves, leather gloves and vinyl gloves. Grip strength was measured using a standard Jamar hand dynamometer. A pegboard task was used to measure the dexterity of the subjects. Manipulability was measured using knot-tying task. Models were developed with linear modeling techniques. Hand breadth was found to be the most contributing factor for all the three tasks.

1 Introduction

The United States, next to China and India, is the world's third most populous place. Reports from the US Department of Statistics (USDS) indicate that the population is highly diversified with people of following origins, i.e., Caucasians, Hispanics, African Americans, Asians and others. These reports identify non Caucasians as Minority population employed mostly as operators, technicians and in other blue collar work. This has had its ramifications on the workforce. Studies (Imrhan et al., 1993; Imrhan and Younes, 1996; Okunribido and Olajire, 1999; and Pennathur and Downing, 2003) reveal that the anthropometrics of the people are race, gender and age dependent. The common finding of these studies is that the hand dimensions of Southeast Asian people are smaller than their Western counterparts. Okunribido and Olajire (1999) measured the hand dimensions of Nigerian farm women and compared with American and European females. They identified that American and European females had narrow and thinner hands than the Nigerian women. In an effort to identify the relation between anthropometry and productivity, Bishu and Mishmash (2001) studied the anthropometric differences between populations in identical manufacturing plants in the United States and Tijuana (Mexico). They found a ten percent difference in the anthropometrics between the two populations. This difference was causing more turnover in Mexico facility.

* Corresponding author.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 493–501, 2007.
© Springer-Verlag Berlin Heidelberg 2007

While differences in anthropometric dimensions among minority populations are well published, not many studies have addressed the issue of effect that these differences have on performance. In this regard, strength performances of minority populations have been studied by a number of researchers (Imrhan et al., 2005; Wang et al., 2005; Eksioglu et al., 1996; Nazaruddin, 2005; and Kim et al., 1992). Our objectives were three fold:

- To perform a collective anthropometric evaluation of the different ethnic groups in the United States. The secondary objective was to compare the hand dimensions of the minority population with the available hand anthropometric data of the US population.
- To valuate performance differences among these minority populations,
- And to develop predictive models relating performance to upper arm dimensions.

While objectives 1 and 2 have been reported elsewhere, this report pertains to the third objective. Models to predict hand dexterity, manipulability and strength performances based on anthropometry have been developed.

2 Method

Data was obtained from an unpublished master's theses (Gnaneswaran, 2005). The data involved checking hand performance on a host of measures for four different minority populations. Fifteen participants each, from Hispanics, African Americans, Vietnamese Americans and Asian Indians, participated in an experiment. Performance measures included grip strength, dexterity, and manipulability. Selected anthropometric measures of the upper extremity were recorded. Participants were mostly students from University of Nebraska. The average age of the female subjects was 23 years and that of male subjects was 29 years. Anthropometric data consisted of hand length (HLength), hand breadth (HBreadth), upper-arm length (UArm Length), forearm length (Farm Length) and hand volume.

Linear modeling techniques were used to develop the prediction models. Two types of regression models were tried. Initially full models were developed to determine the best fit. The R^2 values of the models were used to determine the best fit. Unfortunately, while these could discriminate among a number of candidate models and give the best fit, the intercepts were dispersed widely preventing an objective comparison of regression coefficients. Since the main objective was to develop predictive models, and consequently see if the differences in performances among groups of minority populations could be explained by differences in coefficients, regressions through the origin were tried. The rationale being that, while such a regression will give unreliable coefficients of determination ($R^{2,}$), they will produce beta coefficients that could be compared.

3 Results

The R^2 values of the performance models were estimated for the different hand conditions (Table 1).

Table 1. R^2 values of performance measures

Population	Grip Strength	Dexterity	Manipulability
African American	0.6691	0.3507	0.5073
Hispanics	0.6935	0.1487	.3856
Vietnamese	0.6604	0.3158	0.3507
Asian Indian	0.7060	0.1127	0.4046

The polar graphs of the R^2 values showed that the models were different for all the groups and for all hand conditions. Figure 1- Figure 5 shows the polar graphs for the different models based on hand conditions.

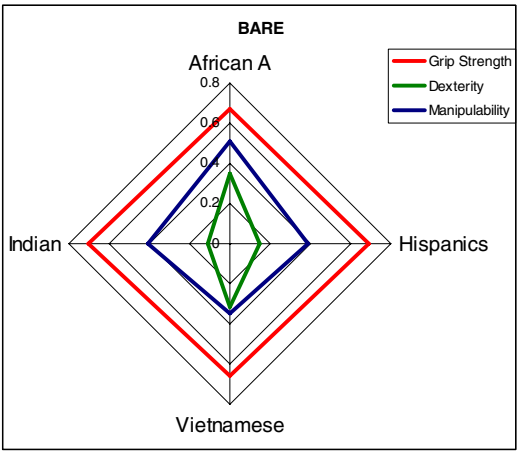


Fig. 1. Polar graph of R^2 values for bare hand condition

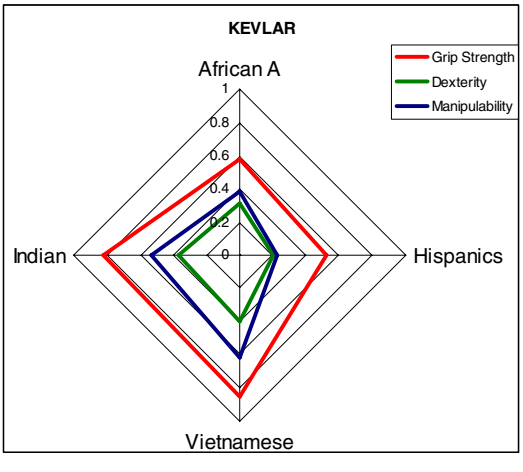


Fig. 2. Polar graph of R^2 values for Kevlar glove condition

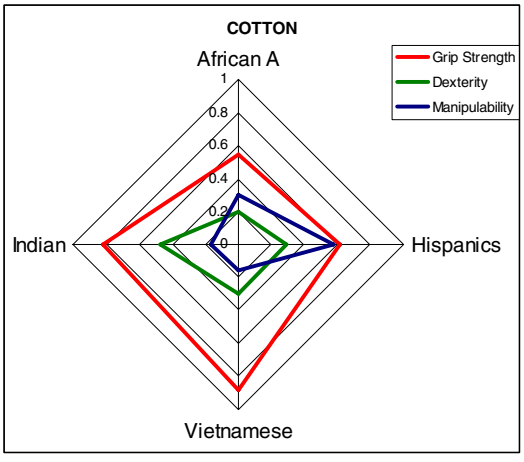


Fig. 3. Polar graph of R^2 values for Cotton glove condition

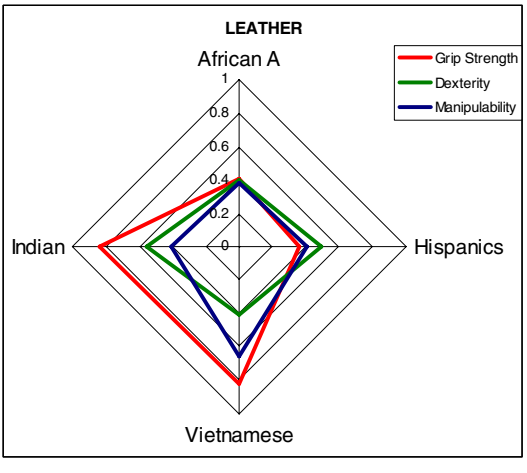


Fig. 4. Polar graph of R^2 values for leather glove condition

The intercepts for the full models were found to be widely scattered (Table 2). To control the effect of intercepts, the regression was forced through the origin. The uncorrected sum of squares was used in this analysis to eliminate negative coefficient of determination.

Table 3 shows the estimated beta coefficients when regression through origin was performed. Figure 6 shows the graphical representation of the predictors of the performance measure for each ethnic group and hand condition.

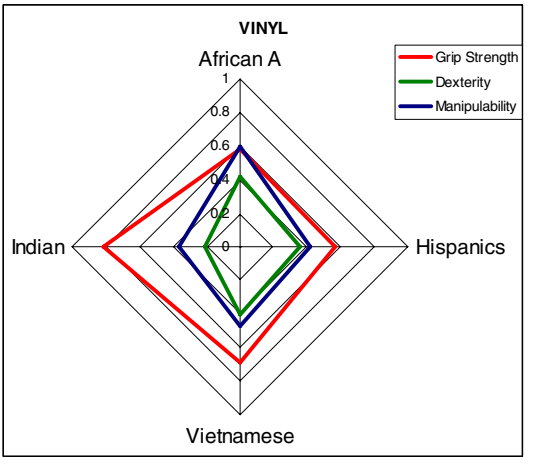


Fig. 5. Polar graph of R^2 values for vinyl glove condition

Table 2. Intercept values for the full model

HAND CONDITION	MINORITY POPULATION	GRIP STRENGTH	DEXTERITY	MANIPULABILITY
BARE	AFRICAN AMERICAN	-4.7693	69.7892	4.5170
	HISPANICS	-121.8664	78.0014	2.6170
	VIETNAMESE	-58.9848	412.2370	2.3083
	ASIAN INDIAN	-61.6157	50.8608	33.9112
COTTON	AFRICAN AMERICAN	34.4612	-41.8162	-49.2726
	HISPANICS	-66.0885	60.7768	39.9473
	VIETNAMESE	-123.7366	608.9459	-129.7367
	ASIAN INDIAN	-159.3398	258.3434	44.9530
KEVLAR	AFRICAN AMERICAN	-87.3063	84.7259	-18.8320
	HISPANICS	-146.5332	61.6350	49.7580
	VIETNAMESE	-135.5088	544.7400	-2.1802
	ASIAN INDIAN	-51.7884	111.1654	-59.3495
LEATHER	AFRICAN AMERICAN	-54.9239	-276.1522	22.0065
	HISPANICS	-58.0842	417.5447	-79.2424
	VIETNAMESE	-98.1584	593.1447	-172.8401
	ASIAN INDIAN	-25.6132	85.6781	-60.8149
VINYL	AFRICAN AMERICAN	-66.9010	36.5674	6.8257
	HISPANICS	-200.2939	189.7316	1.0035
	VIETNAMESE	-107.9593	541.9359	-2.3540
	ASIAN INDIAN	-29.2733	75.8571	20.3690

Table 3. Anthropometric coefficients for the different models

BARE						
Performance Measures	**Minority Population**	**HLength**	**HBreadth**	**UarmLength**	**FarmLength**	**Size**
Grip Strength	African American	0.3703	5.1447	-2.3258	0.9983	0.0776

Table 3. (*Continued*)

	Hispanics	-3.8119	4.6845	-0.0336	1.9118	0.0390
	Vietnamese	1.0770	1.5743	0.3568	-1.6965	0.1025
	Asian Indian	-13.5843	5.0646	2.6553	5.4793	0.0270
Dexterity	African American	0.4883	2.4969	-0.9778	2.4753	0.0073
	Hispanics	0.6765	-7.1103	0.6238	4.4107	0.0395
	Vietnamese	-16.9318	2.9359	-1.6621	21.3834	-0.2824
	Asian Indian	1.1595	2.9706	0.4061	-0.0678	-0.0106
Manipulability	African American	-0.4226	0.2343	0.9777	-0.4852	-0.0078
	Hispanics	-1.7389	0.6745	0.8657	-0.0340	0.0280
	Vietnamese	0.7841	-0.7401	0.3951	-0.3137	-0.0020
	Asian Indian	1.5077	1.5756	-0.8033	-0.1622	-0.0072
COTTON						
Grip Strength	African American	0.9688	3.4500	-0.8578	0.5964	-0.0332
	Hispanics	1.3918	-5.4048	2.4324	1.2956	-0.1233
	Vietnamese	-1.2648	4.9985	0.6581	0.4132	-0.0743
	Asian Indian	-3.5864	3.6556	0.3005	2.2275	-0.0240
Dexterity	African American	-0.4648	5.5223	0.4298	0.0984	0.0465
	Hispanics	-2.0755	-15.4850	2.7987	10.1602	-0.1075
	Vietnamese	-13.8948	-1.2629	-2.5331	14.8604	0.2954
	Asian Indian	3.3999	4.3893	-0.4221	-1.8586	0.0585
Manipulability	African American	0.0326	-0.0671	0.3844	-0.0401	0.0148
	Hispanics	-3.7995	8.9466	2.7773	-4.1482	0.0092
	Vietnamese	-3.4095	0.5428	-1.6596	4.2685	0.0763
	Asian Indian	-4.8196	3.9694	0.9786	1.3141	0.0190
KEVLAR						
Grip Strength	African American	-0.3165	5.3357	-0.5246	0.1185	-0.0204
	Hispanics	0.1021	0.3991	-0.6229	2.1071	-0.0279
	Vietnamese	-6.8978	7.5108	-0.2976	3.9167	-0.0234
	Asian Indian	-14.7541	7.6983	3.0714	4.5581	0.0358
Dexterity	African American	-1.6788	6.2400	0.4010	0.9957	0.0319
	Hispanics	-1.1756	-4.6633	0.0393	6.1845	0.0176
	Vietnamese	-2.9148	-11.0899	-0.1799	10.6025	0.1495
	Asian Indian	0.8572	1.1204	0.7585	0.5781	0.0039
Manipulability	African American	-0.3589	0.3275	0.4551	-0.0604	0.0291
	Hispanics	-0.1805	2.3899	-0.0683	-0.3272	0.0099
	Vietnamese	-0.1384	2.2236	1.0820	-1.8117	0.0151
	Asian Indian	2.3924	1.8773	-0.6339	-1.0116	-0.0117

Table 3. (*Continued*)

LEATHER						
Grip Strength	African American	-0.1605	4.7955	-1.2260	0.8342	-0.0090
	Hispanics	-0.5561	1.4263	-0.3945	1.8074	-0.0262
	Vietnamese	-3.8955	6.8828	-0.4892	2.0004	-0.0314
	Asian Indian	-6.7891	3.2789	1.4240	3.2206	-0.0473
Dexterity	African American	-1.3711	9.7440	-4.7767	8.9318	-0.0877
	Hispanics	-0.1962	-30.1606	-1.5382	23.4084	-0.1315
	Vietnamese	1.3381	-4.6103	-0.6169	3.1888	0.2935
	Asian Indian	1.9281	11.2627	0.1179	-4.9701	0.1299
Manipulability	African American	0.8945	2.9032	-2.1905	2.3188	-0.0264
	Hispanics	-9.3005	13.8005	4.8788	-5.2910	0.1072
	Vietnamese	-2.0342	10.7261	2.1793	-5.2509	0.0281
	Asian Indian	8.7669	9.4634	-5.1498	-3.4303	0.1078
VINYL						
Grip Strength	African American	0.3375	3.1155	-0.3729	0.2216	0.0176
	Hispanics	-1.8221	7.8995	-1.2316	0.9694	-0.0019
	Vietnamese	-9.3685	6.4141	-0.9483	7.2287	-0.0023
	Asian Indian	-10.3778	7.7383	0.5401	4.9530	0.0082
Dexterity	African American	0.0031	-0.2947	0.7220	1.6563	0.0311
	Hispanics	-1.4327	-12.1983	2.7949	6.2440	-0.0132
	Vietnamese	6.0565	-9.7119	1.1735	1.7597	0.0195
	Asian Indian	3.5296	3.7225	-1.3941	0.1684	-0.0011
Manipulability	African American	-0.6201	-2.1688	1.5722	-0.2341	0.0145
	Hispanics	-0.9257	2.1173	0.5075	-0.5359	0.0038
	Vietnamese	2.3001	-0.2073	-0.1482	-1.3011	0.0138
	Asian Indian	-0.8688	1.5194	0.4274	-0.0977	0.0017

Comparing the coefficients it is seen hand breadth to be a better predictor of performance for every group and for different hand conditions. However, forearm length and hand length were also found to be better predictors of grip strength for the Asian Indians and Vietnamese. Dexterity of the Hispanics and the Vietnamese was also found to be predicted by their forearm length.

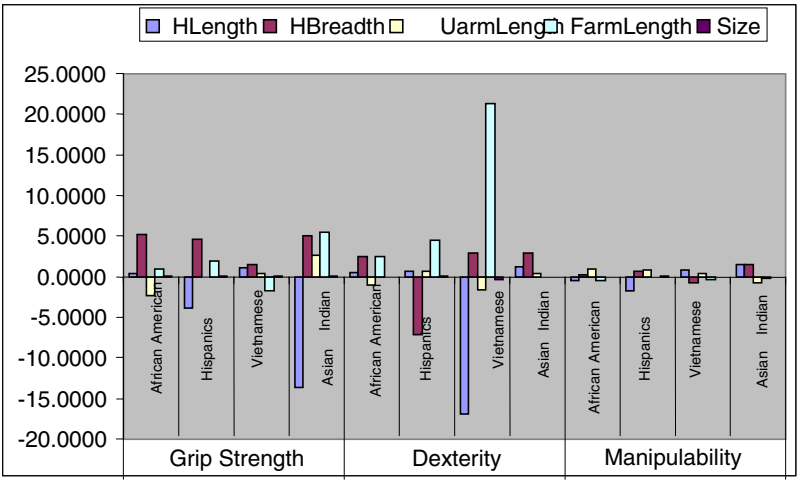


Fig. 6. Anthropometric predictors of performance for bare hand condition

4 Conclusion

The results of this study showed that anthropometry had a significant effect in the performance levels of the four ethnic groups. Given the necessary hand dimensions, the proposed regression models can be used to predict performance levels of workers in industries where experimental evaluations are unrealistic.

References

1. Bellinger, T.A., Slocum, A.C.: Effect of protective gloves on hand movement: An exploratory Study. Applied Ergonomics 24(4), 244–250 (1993)
2. Bishu, R.R., Mishmash, M., Kim, B.: Anthropometric Differences in U.S. and Tijuana, Mexico Populations. In: Bittner, A.C., Champney, P.C., Morrissey, S.J. (eds.) Advances in Occupational Ergonomics and Safety, pp. 507–514. IOS Press, Amsterdam (2001)
3. Chang, C.H., Wang, M.J.J., Lin, S.C.: Evaluating the Effects of Wearing Gloves and Wrist Support on Hand-Arm Response While Operating an In-Line Pneumatic Screwdriver. International Journal of Industrial Ergonomics 24(5), 473–481 (1999)
4. Cochran, D.J., Albin, T.J., Bishu, R.R., Riley, M.W.: An analysis of grasp force degradation with commercially available gloves. In: Proceedings of the 30th Annual Meeting of the Human Factors Society, pp. 852–855 (1986)
5. Eksioglu, M., Fernandez, J.E., Twomey, J.M.: Predicting peak pinch strength: Artificial neural networks vs. regression. International Journal of Industrial Ergonomics 18, 431–441 (1996)
6. Griffin, D.R.: Manual Dexterity of Men Wearing Gloves and Mittens, Fatigue Lab, Harvard University, Report No. 22 (1944)
7. Imrhan, S.N., Mandahawi, N., Sarder, M.D.: Regression Models for Predicting Hand Grip Strength Across A Wide Age Range. In: Proceedings of the 10th Annual International Conference on Industrial Engineering – Theory, Applications and Practice, pp. 570–573 (2005)

8. Imrhan, S.N., Nguyen, M.T., Nguyen, N.N.: Hand Anthropometry of Americans of Vietnamese Origin. International Journal of Industrial Ergonomics 12, 281–287 (1993)
9. Imrhan, S.N., Younes, S.: Comparison of Anthropometric Ratios across Populations. In: Mital, A., Krueger, H., Kumar, S., Menozzi, M., Fernandez, J. (eds.) Advances in Occupational Ergonomics and Safety I, pp. 66–70 (1996)
10. Kim, C.H., Marley, R.J., Fernandez, J.E.: Prediction Models of Grip Strength at varying wrist positions. In: Kumar, S.(ed.) Advance in Industrial & Ergonomics Safety IV. Taylor & Francis (1992)
11. Larsen, L.J.: The Foreign-Born Population in the United States: 2003. U.S. Department of Commerce Economics and Statistics Administration (2004)
12. McKinnon, J.: The Black Population in the United States: March 2002. U.S. Department of Commerce Economics and Statistics Administration (2003)
13. Muralidhar, A., Bishu, R.R.: Glove Evaluation: A Lesson from Impaired Hand Testing. In: Aghazadeh, F. (ed.) Advances in Industrial Ergonomics and Safety VI. Taylor & Francis (1994)
14. Nazaruddin, Z.T.: Grip strength prediction for Malaysian industrial workers using artificial neural networks. International Journal of Industrial Ergonomics 35, 807–816 (2005)
15. Okunribido, O.O., Olajire, K.A.: A Survey of Hand Anthropometry of Female Rural Workers in Ibadan, Western Nigeria. Ergonomics SA 11, 2–6 (1999)
16. Pennathur, A., Dowling, W.: Effect of Age on Functional Anthropometry of Older Mexican American Adults: A Cross-Sectional Study. International Journal of Industrial Ergonomics 32, 39–49 (2003)
17. Rajulu, S.L., Klute, G.K.: Anthropometric survey of the astronaut applicants and astronauts from 1985 to 1991. UNL Libraries Microform (1993)
18. Ramirez, R.R., de la Cruz, G.P.: The Hispanic Population in the United States: March 2002. U.S. Department of Commerce Economics and Statistics Administration (2003)
19. Reeves, T., Bennett, C.: The Asian Population in the United States: March 2002. U.S. Department of Commerce Economics and Statistics Administration (2003)
20. Wang, M.J., Bishu, R.R., Rodgers, S.H.: Grip Strength Changes When Wearing Three Types of Gloves. In: Proceedings of the Fifth Symposium on Human Factors and Industrial Design in Consumer Products, Interface 87, Rochester, NY (1987)

Investigating the Differences in Web Browsing Behaviour of Chinese and European Users Using Mouse Tracking

Lee Griffiths and Zhongming Chen

The School of Computing, Science and Engineering, University of Salford
l.s.griffiths@salford.ac.uk, z.m.chen@pgt.salford.ac.uk

Abstract. The World Wide Web has become a ubiquitous information source and communication channel. With such an extensive user population, it is imperative to understand how users view Web pages. Studies of Web browsing behaviour aimed at different cultures have previously been carried out using methodologies such as questionnaires, observation and expensive eye-tracking. Mouse-tracking however, has not been previously widely applied to studies of Web browsing behaviour. This paper presents an exploratory study in which Web browsing behaviour was investigated with a help of a remote proxy mouse tracker. Furthermore, this paper compares the browsing behaviour of European users with Chinese users. This comparative study tries to explore whether or not there exists any differences in expected menu positions between Chinese and European users using mouse tracking methodology.

Keywords: Cross-culture, Eye-mouse Correlation, Mouse Track Patterns.

1 Introduction

Do users in different countries, with different ethnic origins browse Web pages based on different behaviour patterns? With the shifting geography of the global Internet population towards Asia and the Far East, Web developers need to understand whether there should be different approaches to Web application design, based on differences in the audience. Chau et al [1] suggest that the online behaviours of consumers are subtlety different in nature from traditional consumer behaviour due to the unique characteristics and interplay of technology and culture. More recently Microsoft® has recognised the need to address localisation in their Visual Studio 2005 suite allowing developers to easily automate culture mapping of Web applications. The Hong Kong and Shanghai Banking Corporation Limited (HSBC) has also long recognised the power of localisation in its marketing and business campaigns. There is no doubt the these organisations are gearing up to an anticipated geographic shift in the worlds economy.

A number of studies have in the past been carried out using a series of questionnaires or specified tasks, to explore differences in browsing behaviour in terms of user perception and satisfaction levels. Through an experiment using a series of questionnaires, Simon [2] indicates that perception and satisfaction differences

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 502–512, 2007.
© Springer-Verlag Berlin Heidelberg 2007

exist between the cultural clusters and gender groups within Asia, Europe, Latin & South America and North America. From a study by Han [3] the effects of alphanumerical display formatting on search time among Chinese and American users were measured with result indicating that the formatting of a list on the screen affected American users significantly with a vertical display format leading to faster responses to tasks. The formatting of a list on the screen had no significant effect on the response time among Chinese users. These and other investigations are hinting towards a need to investigate cultural differences further.

Mouse tracking is certainly a poor man's version of eye-tracking, but it is perhaps more efficacious in that it can be easily incorporated into any Web application to provide valuable user interface usage data. Based on the evidence provided by previous studies [4, 5, 6] of eye-mouse movement correlation, it should be effective to evaluate user perception using a mouse tracking methodology. The work presented here attempted to measure the difference in search performance between Chinese and European user with regards to location of Website menus. Based on the work by McCarthy et al [7], a prediction was made that the participants of the two groups would have the same search performance times with a left-justified menu bar followed closely by the menu bar located at the top and right of the page, and finally by the bottom-justified menu bar. In addition, this study derived three mouse movement patterns after careful analysis of 320 mouse track samples.

2 Experiment Design

2.1 Mouse Tracker

The experiment was designed to collect mouse information by a proxy-based mouse tracker that could be accessed through any Web browser without altering user's browsing habits and experience. This was to ensure that the automated collection of mouse trace data was neither perceptible by the user nor intrusive. Furthermore, to prevent any delays due to network traffic, experiment pages were only displayed after all data associated with their traffic had been loaded. The use of a proxy-based mouse tracker allowed the authors to carry out the experiment without having to install the developed mouse tracking software on the experimental machines. The authors do recognise that this approach could allow test subjects to perhaps carry out the experimental tasks in an arbitrary manner however the tasks were designed to attempt to reduce any effect caused by this mode of behaviour.

2.2 Experiment Pages

All pages of the experiment belong to complex category [7] that is they consisted of various components such as a search box, images, email alert function and informative content, etc. All pages have a similar layout [7] which comprises of five generally accepted Web layout regions as shown in Fig. 1 below:

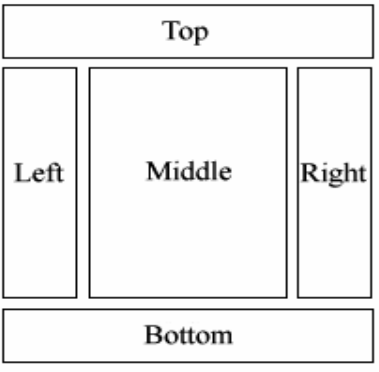


Fig. 1. This shows the layout of all pages in experiment

All participants (*N=40*) were tested on this kind of page layout where the menu bars were located at the top, left, right and bottom side of the experiment page. In order to examine the overall effect of different menu bars on search performance, the study was conducted using four categories of Websites[1] which belong to different fields of interest respectively. Each of these pages were translated into two language versions providing for the two groups, but the layout remained the same. In addition, careful consideration was given to page design in order to avoid the impact generated by cultural knowledge. For example content relating to brands only experienced locally was either avoided of adjusted to be relevant to each cultural group. Moreover, the pages were designed with randomly allocated left, top, right and bottom menus in order to avoid the impact of attention focus and information density [8,9].

2.3 Tasks

It was predicted that the user perception would be one of the main factors which affected mouse trace patterns. The study also assumed that user perception could be classified into different ranks using various difficulty levels of a search task. In order to distinguish difficulty levels, each page was designed with two tasks whose relative difficulty was defined by the easy of location of the search goal. A difficult task (#2) had only one valid goal on a page whilst an easy task (#1) had four. It was assumed that multiple goal targets (#1) would improve search performance and promote simple fixed pattern mouse traces (*see section 5.1*). Tasks goals and their locations are shown in Table 1.

The first task of each test page involved four valid portals or targets that were allocated somewhere on the four menu bars (left, top, right and bottom) with similar information density; for the second task, there was only one valid portal which was placed on a specific menu bar and was potentially more difficult to locate.

[1] Online movie site, government site, education site and online shopping site.

Table 1. This describes the details of tasks for each experiment page

Page	Level	Scenario Description	Locations
Online Movie	#1	Please try to find the SERIES portal on the test page.	Left, Top, Right,Bottom
	#2	Sign up to an EMAIL ALERT SERVICE for the latest on HBO's critically-acclaimed drama.	Right
Government	#1	Imagine that you are now a citizen in Cardiff. In order to seek a suitable job for yourself, try to find the JOBS portal.	Left, Top, Right, Bottom
	#2	As a citizen in Cardiff, it is necessary to know the relevant detail of the tax system. Please find and click the portal for COUNCIL TAX.	Left
Education	#1	Now you are a student in Indiana University (IU), please find the NEWS & INFORMATION portal on the Webpage.	Left, Top, Right, Bottom
	#2	This is the second time you are going to view the homepage of IU. As a new student, it would be a good choice to learn the history of your own university, please try to find the right portal for IU HISTORY.	Top
Online Shopping	#1	Today, you are going to buy a cheap cooking implement which is displayed in the item clearance section, so for this task, try to find the CLEARANCE portal.	Left, Top, Right, Bottom
	#2	After you have bought the cooking implement, unfortunately it is not as good as you expected, now you want to return it. For this task, please try to find the portal for the RETURNSPOLICY.	Bottom

3 Participants

Twenty users for each group were invited to participate in this experiment. In describing their daily Internet usages, most of participants had used the Internet for more than 2 hours; only 9 European participants used the Internet from 0-2 hours. For the European group, 20 participants consisted of 7 females and 13 males, two of them were over the age of 25 while the remaining volunteers were between 18 to 25 years old. For the Chinese group, 11 males and 9 females, only three of them were over 25 years old.

4 Procedure

In order to ensure high-quality data, all participants were only allowed to visit the specified experiment Web site under controlled conditions. Before the formal experiment, users were asked to complete an online background questionnaire to collect relevant demographic information. Once participants entered into the formal

experiment, they carried out a total of eight search tasks following the instruction on a preceding task description page (see Table 1). Participants were only able to move to the next experiment task page after completing the preceding task by finding and clicking on the valid portal links. This process continued until all the tasks were completed. The participants were provided with a free viewing condition, in that they were told to view the pages as normal, with the opportunity to scroll up and down the page at their leisure although the authors were observant that test users did not stray from the tasks.

5 Results

5.1 Mouse Movement Patterns

Through analysing a total of 320 samples of mouse traces based on *N=40* subjects three significant types of movement patterns were found for both groups of Chinese and European users. The following sections show these three kinds of patterns and include the relevant behavioural characteristics:

Straight Pattern. In this pattern, the mouse movement started with a pause in an initial region, while users were presumably visually browsing other regions of the page. It is surmised that once users spotted the link that they were asked to find, they moved the mouse straight to it, and usually terminated the action with a click (as shown in Fig 2.). Perhaps more importantly, the straight pattern could be further described by measuring the time taken to create the pattern. This is because some users start to move their mouse after an initial pause whilst some do not. There are completely different values for these two situations, pause and no pause, and it was assumed that there was an eye-mouse relationship if the initial suspension time of a straight pattern was below 1 or 2 seconds. Thus if there was an eye-mouse relationship then the user expectation of menu positions could be determined as well.

Fixed Pattern. For most users, it seems that there are some fixed regions for the mouse to stay due to their own style of operating the mouse. In this study, it was found that most users like to keep the mouse on the right-side of the page. This might be explained by the fact that they are used to moving the mouse in this area for the purposes of scrolling or unintentional clicking, whilst, they are looking at other regions. Once the destination link is found, then a straight pattern movement will be followed. Therefore it is difficult to measure user expectations of menu positions using this pattern due to the apparent random movement of the mouse and perhaps lack of relationship between the eye and mouse. However, this pattern is perhaps able to suggest hot regions for the mouse cursor to stay which could be exploited by Web designers.



Fig. 2. Screenshot of straight pattern



Fig. 3. Screenshot of fixed pattern

The dotted ellipse in Fig 3 above shows, for a particular subject, that this user moves the mouse to the right-hand blank region of the page, and is doing some scrolling behaviour (up and down). It is assumed that they are also searching around the page at the same time as the time measurements for this trace would suggest. Ultimately, the given search task is completed with a straight pattern movement which is shown within the dotted rectangle.

Guide Pattern. In this, the most intriguing pattern, users appear to search around instinctively using the mouse cursor as some kind of guide. The data for this kind of pattern reveals a continuous movement of the mouse cursor rather than the interrupted Straight and Fixed patterns above. It is difficult to quantify this trace pattern but visually the data would suggest a relationship between mouse and eye movement, although the authors intend to investigate this further.

The trace reproduced in Fig. 4 shows a typical guide pattern sample where it can be seen that the user was searching from the left-side menu to top-side menu using a mouse cursor as a guide shown in the dotted rectangles. Subsequently, the mouse cursor was moved to the right-side blank area following a relatively fixed pattern trace and finally the task was completed with the straight pattern.



Fig. 4. The red rectangles shows a typical guide pattern trace through a menu

In contrast to the fixed pattern, the guide pattern contains very interesting information, which usability researchers can use to examine user expectations towards the position of content, be it menus or other important items. This statement does rely on the acceptance of a correlation between eye and mouse movement.

Fig. 5 below shows the proportion of the guide patterns recorded for the whole experiment (including all eight tasks). Through analysing a total 160 mouse trace samples for Chinese group, 44.38% of samples that contain one or more mouse tracks can be validated to be a guide pattern distinctly; and approximately 35% for the 160 European samples. Furthermore, for all 40 subjects, there were only 5 European participants who carried out all tasks without any evidence of a guide pattern at all.

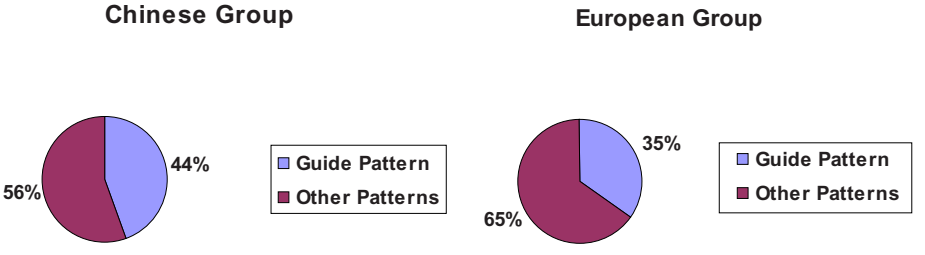


Fig. 5. The proportion of the guide pattern in the whole experiment for both groups

Furthermore, this paper identifies a factor which appears to impact on this kind of variation on guide pattern. Fig. 6 indicates that there exists a variation due to task difficulty level. For example, a difficult to locate target is one where the search goal was in a position which did not meet the user's initial expectation. The effect is apparent for both groups of users and is displayed more frequently with the higher difficulty. Thus there could be a significant correlation between users producing a guide pattern and the location of important objects on the page.

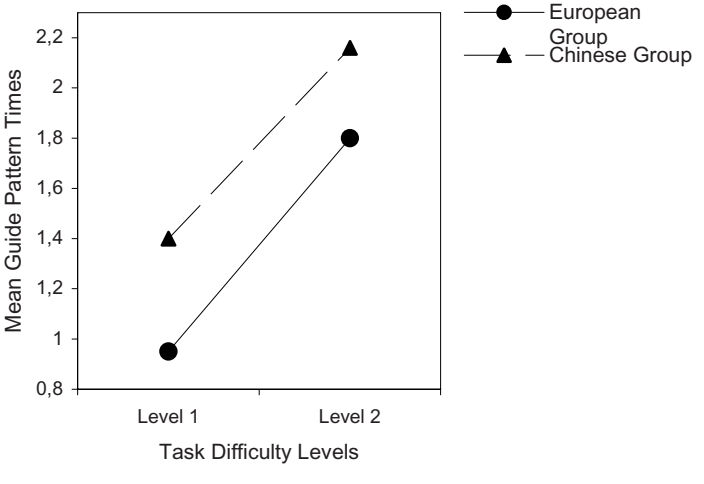


Fig. 6. The variation of proportion of guide pattern

5.2 Differences in Search Performance

In order to examine differences between search performance times, ANOVA was conducted on the time measurements between two groups. The time data collected from the second tasks of each page were analysed and a summary of result are presented in Table 2.

Table 2. ANOVA results on search performance

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	84.1	1	84.1	1.110106	0.293731	3.903367
Columns	2344.85	3	781.6167	10.31721	3.2E-06	2.664109
Interaction	123.35	3	41.11667	0.542733	0.653769	2.664109
Within	11515.3	152	75.75855			
Total	14067.6	159				

Furthermore, table 2 shows the means for each group and each menu position, so that three important results can be concluded as below:

1. There was no significant difference between groups ("Sample") because the calculated F value (1.110106) is less than the critical F value (3.903367), $p > 0.05$.
2. There was a significant difference for both groups (p = 0.0000032) for the effect caused by menu position ("Columns") overall. This shows that the position of the menu significantly affected the search performance times and also indicates that the differences in the means are not by chance alone.
3. It also reveals that there is no significant interaction (p = 0.653769) between groups and menu position.

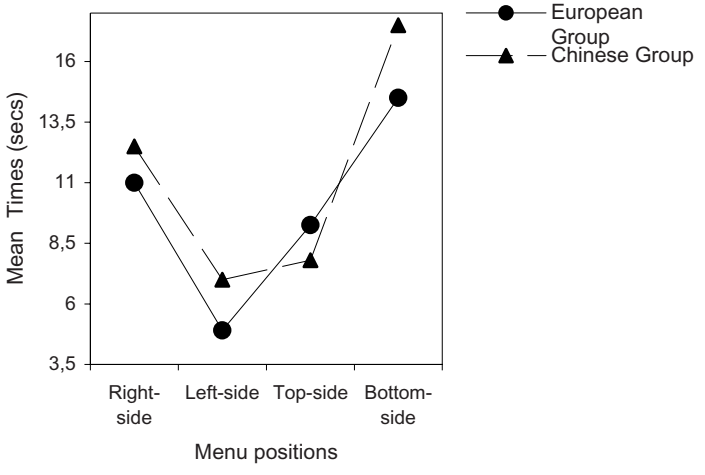


Fig. 7. Interaction effect on search performance times

Lastly from Fig. 7 above it is clear that participants of both groups spent the lowest time on left-side menu of the four. It is much more apparent for the European group than the Chinese group; however Chinese users have a similar search performance on the left-side and top-side menu bar around 7.5sec. As expected, both groups had the worst search performance with the bottom-side menu bar.

6 Conclusion

Based on a careful analysis of our data, this study has found that there are significant differences in the search performance between the different menu bar positions on both groups – indicating that the effect of changing the position of the menu bar was significant. However, it is not driven evidently depending on the region of subjects (the study has eliminated the impact of different cultures). Thus it is inferred that users in different countries would have a similar expectation of menu position.

As expected, this paper confirmed previous works that an apparent relationship exists between gaze position and cursor position on a computer screen during Web browsing, but not always.

This study thus indicates that it could be valuable to categorise mouse trace patterns in a modular way to describe the complexity of regions of a Web page.

This study thus indicates that it could be valuable to categorize mouse trace patterns in a modular way. By combining the analysis of mouse trace patterns and other variables such as time it is possible to evaluate user expectations and perceptions of page layouts.

Acknowledgements. The author would like to take this opportunity to thank all subjects who took part in the experiment.

References

1. Chau, P.Y.K., Cole, M., Massey, A.P., Montoya-Weiss, M., O'Keefe, R.M.: Cultural differences in the online behavior of consumers, vol. 45, pp. 138–143. ACM Press, New York (2002)
2. Simon, S.J.: The impact of culture and gender on Websites: an empirical study, vol. 32, pp. 18–37. ACM Press, New York (Winter 2001)
3. Han, S.: Effects of Alphanumerical Display Formatting on Search Time among Chinese and American Users CHI '06 extended abstracts on Human factors in computing systems
4. Chen, M., Anderson, J.R., Sohn, M.: What Can a Mouse Cursor Tell Us More? In: Correlation of Eye/mouse Movements on Web Browsing. Ext. Abstracts CHI 2001, ACM Press, New York (2001)
5. Mueller, F.: Andrea Lockerd: Cheese: Tracking Mouse Movement Activity on Websites, a Tool for User Modeling
6. Arroyo, E., Selker, T., Wei, W.: Usability Tool for Analysis of Web Designs Using Mouse Tracks, CHI 2006 Work-in-Progress

7. McCarthy, J.D., Sasse, M.A., Riegelsberger, J.: Could I have the Menu Please? An Eye-tracking Study of Design Conventions. In: Proceedings of HCI2003/, 8-12 September 2003, Bath, UK, pp. 401–414 (2003)
8. Rayner, K.: Eye Movements in Reading and Information Processing: 20 Years of Research. Psychological Bulletin 124(3), 372–422 (1998)
9. Granka, L.A., Hembrooke, H.A., Gay, G., Feusner, M.K.: Correlates of Visual Salience and Disconnect: An Eye-tracking Evaluation

The Effect of Morphological Elements on the Icon Recognition in Smart Phones

Chiwu Huang and Chieh-Ming Tsai

Department of Industrial Design and Graduate Institute of Innovation Design
National Taipei University of Technology
1 Chung-Hsiao E. Rd., Sec. 3, Taipei 106, Taiwan
chiwu@ntut.edu.tw

Abstract. This study aims to explore the effect of morphological elements on the icon recognition in smart phone. 42 icons were first selected and classified in a morphological chart based on its visual design elements. Then, icons were evaluated by a group of respondents with or without design background through e-mail. Main findings include: 1) Some morphological elements may affect the recognition rates of icons. Icons imitating real objects and using conventional symbols are better in recognition rate. In contrast, some particular symbols may be difficult to recognize. 2) Gaps may still exist between designers and users. The result of this study shows that the un-answered rate of the respondents without design background is significantly higher than the ones with design background. This may suggests that gaps may still exist between designers and users. Therefore, it is recommended that a designer should bear user in mind when designing icons in order to minimize these gaps.

Keywords: Smart phone, Icon, Recognition rate, Morphological elements.

1 Introduction

The functions of mobile phone in the early days were very basic, ranging from making or receiving phone calls to sending or receiving messages. Nowadays, a much more advanced Smart Phone can deal with almost everything in one's daily lives. A Smart Phone is a mobile phone that incorporates PDA functions in a small handset. Therefore, it is a challenge for engineers to keep so many functions in small space. As a result, the design of the man-machine interface becomes very important. Under the effort of R&D team, the function of many Smart Phones, in terms of both hardware and software, are getting similar. However, differences exist in their interface design, especially the way of interaction and graphic icon design. Icon design can be considered as the entrance to the interaction of man-machine interface. Whether a user can recognize its meaning or not becomes the first step to interact with a phone set.

In an anecdotal observation, however, the recognition rate of icon design in smart phones was not very high. According to Norman's mental models [1], ideally, the design model should be identical with the user's model when designing icons. It leads to a hypothesis that the difference of the recognition rate to an icon design between

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 513–522, 2007.
© Springer-Verlag Berlin Heidelberg 2007

users and designers should be minor. Moreover, in a preliminary study it was suspected that the morphological structure of an icon design might have effects on the recognition of an icon. It leads to this study to explore:

1. Whether there is any difference existing between the recognition rates on icon designs of smart phones between users with and without design background.
2. Whether there is any effect of the morphological elements in icon designs of smart phone on the recognition rate to users.

2 Literature Review

2.1 The Definition of Smart Phone

Smart phone was first coined by Motorola [2]. At the beginning, any mobile phone equipped with functions more than making a phone call was called “Smart phone”. Nowadays, this term becomes much more general. It is not used exclusively by Motorola anymore. The first smartphone was called Simon designed by IBM in 1992 and shown as a concept product that year at COMDEX, the communications industry trade show held in Las Vegas, Nevada. It was released to the public in 1993 and sold by BellSouth. Besides a mobile phone, it also contained a calendar, address book, world clock, calculator, note pad, e-mail, and games. It had no physical buttons to dial with. Instead customers used a touch-screen to select phone numbers with a finger or create facsimiles and memos with an optional stylus. Text was entered with a unique on-screen "predictive" keyboard. By today's standards, the Simon would be a fairly low-end mobile phone [3]. This study summarizes definitions for a smart phone as shown in Table 1.

Table 1. The summary of definitions for a smart phone

appearance	The size is small, short, light and thin. The appearance is not restricted to either mobile phone or PDA.
functions	It is used mainly in voice communication. Digital transmission is also included. It uses advanced mobile operation system. It is equipped with personal information management function. The information can be exchanged or synchronized with other information products.
Input methods	It is not restricted to keyboard or touch sensitive panel. Voice recognition is also possible.

2.2 The Classification of Graphic Icons

At present, graphic icons are widely applied to many areas. The form is not limited to a certain kind of style. Peirce classified signs into three categories, icon, index and symbol [4]. Every category relates differently to its referent:

Icon. Icon means to imitate the physical image of its referent. It is much more concrete. Icon features the characteristics of its subject.

Index. Index relates indirectly to the concept of its referent. For example, a trapezium icon emitting sound weaves means a "loudspeaker".

Symbol. Symbol neither looks alike nor relates to its referent. It is customary used or agreed by most people for a long time.

Due to the advance of computer technology, graphic icons are developed from two-dimensional image into three-dimensional image. Technologies, such as voice recognition, multimedia and animation, are used in Smart Phones. The choice of interaction between man and machine becomes much more abundant. However, the acceptance of users regarding the interface design depends largely on the evaluation of the icon designs.

2.3 Studies of the Evaluation of Icons

The paired-comparison method [5] is most often used in the evaluation of icon design. Respondents are asked to pick up the most fitted name from a list to match the icon. The recognition rate can be determined through statistical analysis. The icon with bad recognition rate is subjected to redesign by the designer according to the opinion from respondents. However, objective method for the analysis of icon design is lacked. Huang [6] uses statistic to analyze the relationship between icon's image vocabulary and its morphological elements for mobile phone. Morphological analysis is a systematic method that breaks down an icon into elements. The elements are classified into items and categories.

The morphological analysis is used in the study to explore the relationship between recognition and morphological elements. The characteristics of an icon with both high and low recognition rates can then be identified.

Chen [7] and Lee [8] have done researches on the evaluation of icon design for mobile electronic products. However, the respondents were divided into groups according to their needs and habits. The respondents are categorized into people with or without design background. The purpose is to identify the difference between two groups on the recognition of icon. The result can be applied to icon design to avoid the gap between designers and users.

2.4 Morphological Analysis

Morphological analysis was originally used in engineering design to develop solutions for problems [9]. Solutions are classified into main categories and sub-categories first. Varies solutions are then proposed through the combination of sub-categories. In this way, solutions can be explored systematically. This method was adopted for the study.

3 Method

3.1 The Research Process

The research process is divided into following five steps:

The Selection of Sample for the Evaluation. 6 smart phones from the top-three operation system, i.e. Symbian, Palm and Windows Mobile Smartphone, were selected

for the study. Six phones were selected, including Nokia 9500, Nokia 7710, Nokia 7650, Siemens SX1, Treo 650 and Dopod 575.

The Collection of Icons from Smart Phones. The graphic icons of top-three best sellers of smart phones were chosen for study. Icons of main functions were selected. As a result, 167 icons were collected.

The Preliminary Analysis of Icons. Icons were categorized into items and categories according to the morphological analysis (as shown in Table 2) for mobile phones proposed by Hwang [3].

The Selection of Icons for Evaluation. 167 icons were screened down into 42 icons. Two steps are employed as below:

Questionnaires. 20 respondents were asked to rank 167 icons in 5-point Likert scales, ranging from 1(used hardly) to 5 (used very often). The result was evaluated by the focus group in the following step.

Focus group. 5 experts in Industrial Design or Visual Communication were invited to form a focus group. They were asked to screen the icons according to the following rules:

- Icons with lower points that have the same function were screen out.
- Icons with similar morphological elements but in poor design quality were screened out.

The Evaluation of Icons. The confusing matrix method [10] was used for the evaluation. First, 42 icons were listed randomly on the top of the computer screen. The meanings were listed randomly on the pull-down menu for the respondent to choose from. A proper meaning was put into the box by the respondent. The result was constructed into a confusing matrix for a further analysis. Icons that matched the intended meaning would have lower confusing rate. In contrast, if icons do not match the intended meaning the confusing rate would be high. Recognition rate was then calculated with equation (1). 66.7% recognition rate was recommended by ISO [11]. Through the analysis of confusion matrix, the recognition rate and reasons for misinterpretation can be determined. Besides, independent sampling t test was conducted to determine whether the design background of the respondent affected the recognition rate.

$$\text{(Number of correct choices / Number of respondents)} \times 100\% = \text{Recognition rate}. \quad (1)$$

3.2 Participants

Due to the availability and willingness of participants, convenient sampling was adopted for the study. 80 young students and office workers were approached by e-mail as participants. Among them, 40 participants have design background whereas the other 40 do not have design background. The questionnaires were distributed through e-mail.

Table 2. An example of morphological chart

Categories	Sub-categories		
A. Types			
	A1. icon	A2. index	A3. symbol
B. Styles			
	B1.orthographic image	B2.diagonal image	B3.3-D image
C. Auxiliary elements			
	C1. yes	C2. no	
D. Backgrounds			
	D1. yes	D2. no	
E. Ways of presentations			
	E1. whole image	E2. partial image	

4 Results and Discussions

4.1 Basic Data

80 questionnaires were distributed through e-mail services. Consequently, 68 responds were received. 9 of them were rejected for missing or repeating values. Therefore, 59 effective responds were collected. The responding rate was 73.8%.

Five (8.5%) respondents are under 19 years old. 19 (32.2%) respondents are 20-29 years old. 23 (39.0%) respondents are 30-39 years old. 9 (15.3%) respondents are 40-49 years old. 3 (5.1%) respondents are over 50 years old.

Regarding background, 31 (52.5%) respondents have design background whereas 28 (47.5%) do not have design background.

4.2 Icons Recognition Rate Analysis

28 icons meet ISO standard with the recognition rate over 66.7% while the rest of 14 icons, i.e. one third of tested icons, fail to meet the standard. (See Table 3) This result conforms to the anecdotal observation mentioned in the introduction.

Table 3. The recognition rates of icons

. No.	A1	A2	A3	A4	
Address book					
Recognition rate (%)	81.4	78.0	52.5*	47.5*	
No.	B1	B2	B3	B4	B5
message					

Table 3. (*Continued*)

Recognition rate (%)	98.3	81.4	87.8	69.5	55.9*
No.	C1	C2	C3	C4	
Phone call record					
Recognition rate (%)	76.3	23.7*	22.0*	57.6*	
No.	D1	D2	D3	D4	
setting					
Recognition rate (%)	94.9	94.9	93.2	18.6*	
No.	E1	E2	E3	E4	
camera					
Recognition rate (%)	100.0	100.0	98.3	100.0	
No.	F1	F2	F3	F4	
Synchronize data					
Recognition rate (%)	62.7*	57.6*	52.5*	72.9	
No.	G1	G2	G3	G4	
calendar					
Recognition rate (%)	98.3	62.7*	93.2	86.4	
No.	H1	H2	H3	H4	H5
Media player					
Recognition rate (%)	100.0	98.3	100.0	59.3*	89.8
No.	I1	I2	I3	I4	
Check list					
Recognition rate (%)	72.9	66.1*	72.9	49.2*	
No.	J1	J2	J3	J4	
Note book					
Recognition rate (%)	94.9	69.5	67.8	69.5	

* Recognition rate < 66.7%

4.3 Analysis of the Recognition Rate Against Design Background

Table 4 shows the recognition rate, error rate and un-answered rate for the respondents with design background are 77.8%, 17.9% and 4.3% respectively, whereas, the respondents without design background are 70.7%, 18.9% and 9.3% respectively. It is observed that the recognition rate for the respondents with design background is higher than the respondents without design background. However, both the errors rate and un-answered rate for the respondents without design background are higher than the respondents with design background. Furthermore, the data in Table 3 were tested in independent sampling t test. The result showed that the differences between two groups in both recognition rate (p 0.05) and errors rate (p 0.05) are insignificant. However, the unanswered rate of the respondents without design background is significantly higher than the ones with design background (p 0.05).

It can be inferred from the result above that the respondents with design background may be better in recognizing the meaning of the icons than the ones without design background because they are well-trained in visual design. On the

contrary, the respondents without design background may have difficulties in recognizing the icons for the lack of design training. Therefore, it is recommended that a designer should bear users in mind when designing icons. Characteristics of different users should be considered. To use straightforward designs as much as possible may be helpful. Besides, a user evaluation could be very important for improving the design quality. Moreover, recognition errors may happen if the designer designs icons based solely on his own experience.

Table 4. The average of recognition rate, error rate and un-answered rate

respondents	Number	Recognition rate s. d.	Error rate s. d.	Un-answered rate s. d.
With design background	31	77.8 (22.0)	17.9 (18.8)	4.3 (6.9)
Without design background	28	70.7 (23.7)	19.9 (18.1)	9.3 (10.1)
total	59	74.2 (23.0)	18.9 (18.4)	6.8 (9.0)

4.4 The Effect of Morphological Elements

The relationship between the morphological elements of icons and their recognition rate is discussed in this section to determine the morphological effects on icon recognition.

Types. The types of morphological elements may affect the recognition of icons.

Icons are better in expressing meaning. An icon depicting a real object is much better in expressing its meaning and will have a higher recognition rate. For example, icons with the highest recognition rate in the study, E1, E2, E3 and E4 use an image of camera to represent the function of taking a picture. Respondents can easily recognize its function is related to a camera. Other examples, B1, G1 and J1 use the image of envelop, calendar and notebook respectively to represent their functions. They bear also very high recognition rate.

Conventional Symbols Have Higher Recognition Rate. A conventional symbol may have no direct connection to its meaning. However, once people get use to it, however, its meaning can also be accepted. For example, H1, H2 and H3 use a music note to represent playing music. H5 uses a triangle shape to represent "playing" just like the playing symbol in our disc player. D1, D2 and D3 belong to index category. They use a wrench to symbolize "setting-up" something. They all have higher recognition rate than their counterpart.

Particular Symbols May Be Difficult to Recognize. Symbols for particular usages may be difficult to recognize. For example, H4 means "Real Player", media player software in PC. However, not every one was familiar with it. Its recognition rate is only 59.3%.

Auxiliary Elements. Auxiliary elements may be helpful in recognizing the icon. For example, A1 has a telephone handset on a book which means "address book". The meaning seems to be obvious to most respondents. A2 is an image of a partial note book with a telephone icon on it. The meaning is also obvious. The icon of telephone may be helpful in recognizing it. In addition, G1 and G3 use "number" to address that they are calendars.

However, the auxiliary elements may also cause difficulty in recognize meaning if it is too small. For example, A4 is an index file with a person's photograph on it. However, the photo is so small that respondents might fail to see it. That may account for the low recognition rate. The same situation may also apply to B4 and I4.

Ways of Presentations. To use of partial image in icon design may cause difficulties in icon recognition. For example, G2, I2 and J2 use partial image of something to signify calendar, check list and notebook respectively. They may be stylish; however, details may be missing so that the meaning is lost. In addition, the effect of icons with background and icons in different style may not be obvious since some of those icons recognition rate are lower than ISO standard, whereas, some of them are quite high.

4.5 The Analysis of Confusing Matrix

Icons with Similar Design that Represent Different Meanings. Icon C2 and C3 mean "phone call records". However, 55.9% and 59.3% respondents respectively recognize them mistakenly as "synchronize data". On the contrary, 23.7% respondents misinterpret icon F2 as "phone call records" instead of "synchronize data". This may be because their designs are too similar for to distinguish them from each other. They all use two arrows to represent data flows which may cause the confusion. Same thing happens with icon F3 that 27.1% respondents mistakenly recognize it similar to icon H5, "media player".

Icons with Different Design but Similar Wording. Icon C4, "phone call records", is mistakenly recognized as "address book" by 37.3% respondents. This may be because the wordings in Chinese for "phone call records" (tong-shuin-ji-lu) and "address book" (tong-shuin-lu) are very similar to each other.

Misleading Icon Design. Icon D4 means "setting", however, its recognition rate is only 18.6%, the lowest among 42 icons. The design is a card with a list on it which seems misleading to recognize its real meaning.

Icons with Wrong Hints. Certain icons might include inadequate hints in their design which may cause confusion. For example, 35.6% respondents recognize icon B5 as "media player" instead of "message". A music note and a photo are put in front of an envelope. Music note may hint that this relates to "media". A pencil is placed on a ring book in icon G2. 17.0% respondents recognized it as "note book" in stead of "calendar". The pencil may hint that this is a "note book".

5 Conclusion

This study explored the relationship between the recognition rate and morphological elements based on the graphic icon design of three top-seller smart phones in Taiwan.

The icons were first categorized according to their morphological elements. The cognition rate was then calculated. Finally, the confusion matrix of icons and meanings were constructed to determine the cause of confusion. The study concludes:

The Effects of Morphological Elements on the Icon Recognition. It includes:

- Icons are better in expressing meaning.
- Conventional symbols have higher recognition rate.
- Particular symbols may be difficult to recognize.

Gaps May Exist between Designers and Users. The result of this study shows that the un-answered rate of the respondents without design background is significantly higher than the ones with design background. This suggests that gaps may exist between designers and users. Therefore, it is recommended that a designer designing icons should bear user in mind in order to minimize these gaps.

Some Recommendations for Icon Design

- "Icons" imitating real objects may be helpful in increasing the recognition rate.
- Conventional symbols have higher recognition rate.
- Proper use of auxiliary elements in icons may be helpful in increasing the recognition rate.
- A whole image may be better than a partial image in terms of recognition.

Causes for Icon Confusion. Causes for icon confusion determined through confusion matrix analysis are list below:

- Similar design
- Similar wording
- Not relevant to the meaning
- Wrong hints

Acknowledgments. This study is sponsored by the National Science Council of Taiwan (NSC95-2221-E-027-027).

References

1. Norman, D.A.: The design of everyday things, pp. 189–191. Basic Books, New York (1988)
2. Tsai, C.M.: The Evaluation of GUI Icons on Smart Phones (in Chinese). A thesis of Graduate Institute of Innovation and Design, National Taipei University of Technology (2006)
3. Wikipedia: Smartphone (2007) http://en.wikipedia.org/wiki/Smartphone#History
4. Fiske, J.: Introduction to Communication Studies (Chinese edition, Interpretation by Chang, C.H.). Yuan-Liou Publishing, Taipei (1995)
5. David, H.A.: The Method of Paired Comparisons, 2nd edn. Charles Griffin & Company, Limited, London (1988)
6. Hwang, P.W.: A Study of Perceptual Image and Preference for Mobile Phone Human/ Machine Interface (in Chinese). A thesis of Graduate school of Commercial Design, Chung Yuan Christian University (2004)

7. Chen, T.H.: Research on The Functional Requirement and Icon Distinguish of Mobile Phone for Users (in Chinese). A thesis of Graduate School of Industrial Design, National Yunlin University of Science & Technology (2003)
8. Lee, Y.K.: An Evaluation Study on Usability of Iconic Interfaces for Information and Communications Products (in Chinese). A thesis of Graduate Institute of Innovation and Design, Nation Taipei University of Technology (2002)
9. Jones, J.C.: Design Methods, 2nd edn. Van Nostrand Reinhold, New York (1992)
10. Lin, R.T., Chuang, M.C.: Semantic Issues on Icon Design in Man-Machine Interface (in Chinese). Industrial Design Quarterly 20(2), 85–93 (1991)
11. Foster, J.J.: Standardizing Public Information Symbols: Proposals for a Simpler Procedure. Information Design Journal 6/2, 161–168 (1990)

Performance Evaluation of the Wheel Navigation Key Used for Mobile Phone and MP3

Hyun-Wook Jung and Jung-Yong Kim

Dept. of Industrial Engineering in Hanyang University, Ansan, Rep. of Korea
{Hyun-Wook Jung,hyungal}@hanmail.net
{Jung-Yong Kim,jungkim}@hanyang.ac.kr

Abstract. The aim of the study was to investigate the usability of wheel-type navigation key of cell phone and MP3 product. An experiment was designed to evaluate the functional benefit of wheel navigation key by using performance test. A questionnaire was also used to examine the personal preference. Eighteen subjects were recruited. In results, a significant difference was found in performance time between wheel-type and button-type product. In general, the difference was more significant as subject's skill level was higher. In questionnaire, different preference depending on the skill level and key type was reported. In conclusion, it was shown that the wheel-type navigation key improved the performance better as the skill level and search requirement became higher. Therefore, the wheel navigation key would be helpful device for users if we could use them selectively in order to speed up the searching task and replacing simple push button task.

Keywords: Usability, Mobile phone, MP3, Wheel Navigation Key, Button Key.

1 Introduction

Recently, wheel navigation key has been employed to design the interface of digital product such as mobile phone and mp3. The wheel navigation key was used to improve the accessibility to information as well as improve the exterior design. However, we do not have much information about the functional benefit of the wheel navigation key. No empirical data were found to show the benefit of wheel navigation key. Therefore, in this study, the real mobile products with wheel navigation key were selected for experiment and the results were compared with traditional button key. This study used the usability testing techniques (Hix and Hartson, 1993; Nielson, 1994; Ziefle, 2002; Kim et al., 2005) including performance measure and questionnaire. The result of experiment and subjective rating scale for satisfaction are expected to help designers optimize the functional keys in various mobile products.

2 Method

2.1 Participants

Eighteen subjects participated in the experiment. They had experience in using both mobile phone and mp3 with wheel navigation key (WNK). Their age was 23.4(± 3.2)

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 523–530, 2007.
© Springer-Verlag Berlin Heidelberg 2007

years. They were grouped into three: 6 novices, 6 intermediate users, and 6 experts. Novice had experience of using the WNK shorter than one month. Expert had experience longer than 6 months. Other subjects were grouped as intermediate users. Every subject had normal eye-sight and no problem in using their fingers.

2.2 Experimental Hypothesis

The null hypothesis was that there was no difference in performance between wheel navigation key and button key.

2.3 Equipments

Two mobile phones with same function and menu structure were selected. The selected models were IM-8500 (Fig. 1) with WNK and IM-8300 (Fig. 2) with button key made by 'S' electronic company. Both models were manufactured and sold in the market about the same time. Two mp3 devices with same specification except the key were selected. Those were model I (Fig. 3) and model E (Fig. 4) with same function and menu design. A stop watch was used to record the performance time and camcorder was used to record the experimental process.



Fig. 1. Model IM-8500 with WNK



Fig. 2. Model IM-8300 with button key



Fig. 3. mp3 model I



Fig. 4. mp3 model E

2.4 Experimental Design

In order to compare the performance of subjects, 3x3x2 mixed factors design was used. Independent variables were a 'skill level' (novice, intermediate, and expert), 'interface type' (WNK and button key), and 'task' (easy, moderate, and difficult). The skill level was a between factor, and the interface type and task were within factors. Dependent variable was 'performance' measured in time (second). In addition, the level of satisfaction (Nielsen and Levy, 1994) was measured and used as a dependent variable. The subjective rating used 7-scale questionnaire.

Table 1. Experimental design for mobile phone (3x3x2 mixed factors design)

		skill level					
		novice		Intermediate user		expert	
		WNK	button key	WNK	button key	WNK	button key
		IM-8500	IM-8300	IM-8500	IM-8300	IM-8500	IM-8300
task	easy	S1*~S6	S1~S6	S7~S12	S7~S12	S13~S18	S13~S18
	mod.	S1~S6	S1~S6	S7~S12	S7~S12	S13~S18	S13~S18
	diff.	S1~S6	S1~S6	S7~S12	S7~S12	S13~S18	S13~S18

*S1: subject number 1

Table 2. Experimental design for mp3 (3x3x2 mixed factors design)

		skill level					
		novice		Intermediate user		expert	
		WNK	button key	WNK	button key	WNK	button key
		model I	model E	model I	model E	model I	model E
task	easy	S1*~S6	S1~S6	S7~S12	S7~S12	S13~S18	S13~S18
	mod.	S1~S6	S1~S6	S7~S12	S7~S12	S13~S18	S13~S18
	diff.	S1~S6	S1~S6	S7~S12	S7~S12	S13~S18	S13~S18

*S1: subject number 1

2.5 Procedure

Subjects were explained about the purpose and procedure of the experiment. A few minutes were given for subject to familiarize with the devices. Three tasks were performed for each device and the order was counter-balanced to reduce the learning effect. For mobile phone, the easy task was "set the alarm for one hour from now". Moderate task was "delete the blue pictures in first three pages". Difficult task was "find the 38th name 'Cha In-Pyo' in the dialing list and modify the phone number of 'Lee Young-Ha'". For mp3, the easy task was "find the song 'on the street' and play". Moderate task was "find the 'Big Ma Ma's 2nd album and play the song 'all by myself'". Difficult task was "find the 'slow love song' category, find the singing group 'crystal box', and find the song 'like a bird'". The time was measured for each task. After finishing all the tasks, a simple questionnaire was used to measure the satisfaction level.

2.6 Analysis

Analysis of Variance (ANOVA) was used to test the null hypothesis. SAS 9.2 was used for statistical analysis. Post –hoc analysis (Duncan test) was used among variables with significant result in ANOVA. The subject rating was also analyzed.

3 Results

3.1 Performance in Mobile Phone

The skill level ($p<0.01$), interface type ($p<0.01$), and task difficulty ($p<0.01$) showed significant main effect on the performance time. Interaction effects were also observed between task and skill level, skill level and interface type (Table 3).

Table 3. ANOVA for mobile phone performance

Source	DF	SS	MS	F	Pr>F
Skill level(A)	2	245.592	122.796	12.63	**0.0071****
Interface type(B)	1	54.000	54.000	15.93	**0.0072****
Task difficulty(C)	2	1683.370	841.685	210.42	**0.0001****
(A x B)	2	24.333	12.167	3.93	**0.0943***
(A x C)	4	47.630	11.907	2.98	**0.0638***
(B x C)	2	37.000	18.500	8.54	**0.0049****
(A x B x C)	4	9.333	2.333	1.08	0.4104

* significant at $p < 0.1$.
** significant at $p < 0.01$.

In Fig. 5, the difference between interface types was observed. In particular, the difference was more apparent in difficult task.

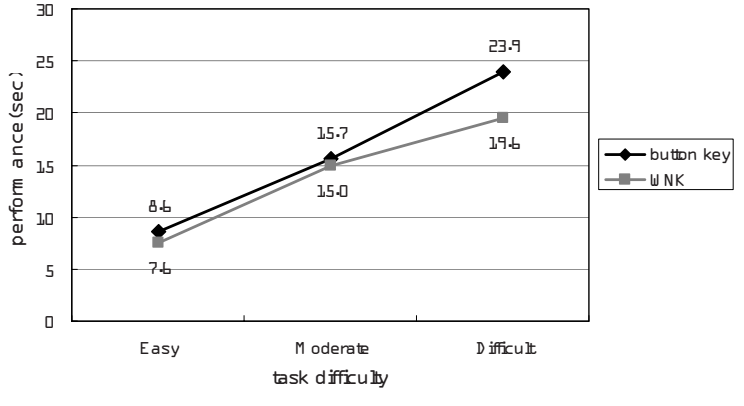


Fig. 5. Performance change depending on the interface type and task difficulty in mobile phone

In Fig. 6, the performance time decreased as the skill level increased as we expected. However the performance difference between interface types was significantly greater among experts than novices.

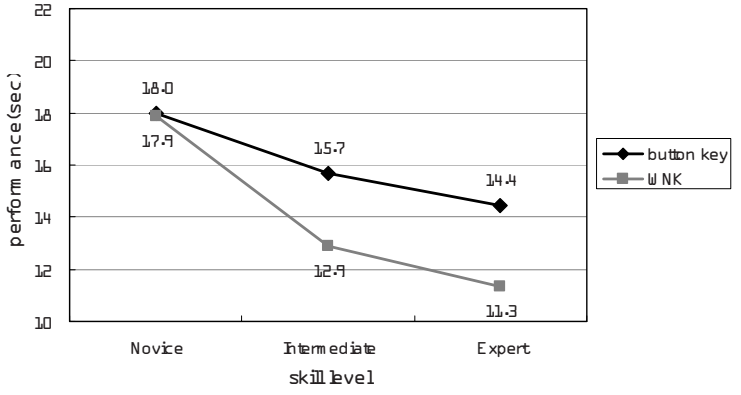


Fig. 6. Performance change depending on the interface type and skill level in mobile phone

3.2 Performance in mp3

The skill level ($p<0.1$), interface type ($p<0.01$), and task difficulty ($p<0.01$) showed significant main effects on the performance time. Interface type and task difficulty showed interaction effect on the performance ($p<0.01$) (Table 4).

Table 4. ANOVA for mp3

Source	DF	SS	MS	F	Pr>F
Skill level(A)	2	216.148	108.074	3.98	**0.0795***
Interface type(B)	1	240.667	240.667	24.47	**0.0026****
Task difficulty(C)	2	480.592	240.296	67.06	**0.0001****
(A x B)	2	5.333	2.667	0.27	0.7713
(A x C)	4	33.074	8.269	2.31	0.1177
(B x C)	2	44.444	22.222	9.88	**0.0029****
(A x B x C)	4	2.556	0.639	0.28	0.8828

* significant at $p < 0.1$.
** significant at $p < 0.01$.

In Fig. 7, the interaction between task difficulty and interface type was shown. The performance improvement with WNK was greater with easy task than difficult task.

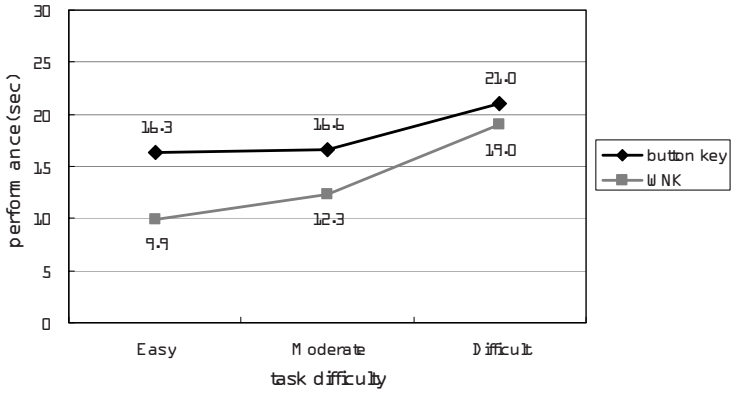


Fig. 7. Performance change depending on the interface type and task difficulty in mp3

Interface type and skill level showed no interaction effect on performance (Fig. 8) although the performance difference between different interfaces was very substantial.

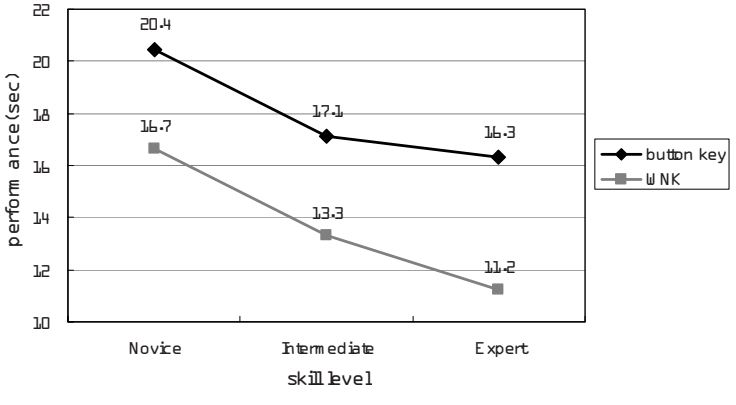


Fig. 8. Performance change depending on the interface type and skill level in mp3

3.3 Subjective Rating for Satisfaction

For mobile phone, novice group preferred button key to the WNK, while expert group preferred WNK to button key (Fig. 9).

For mp3, the general trend in satisfaction level was similar with the mobile phone, but the difference was not as great as the mobile phone (Fig. 10).

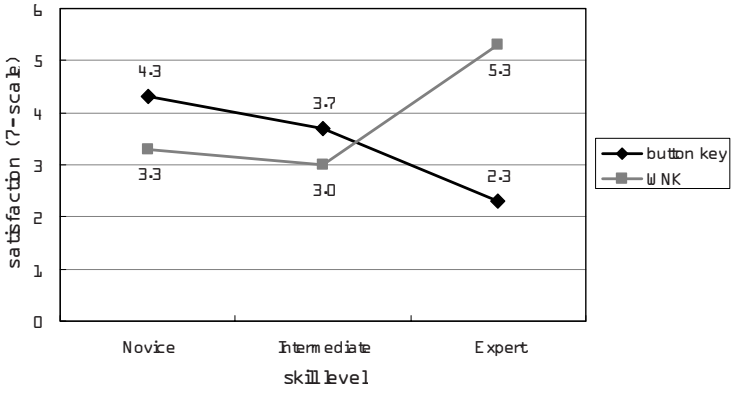


Fig. 9. Subjective rating for satisfaction for mobile phone

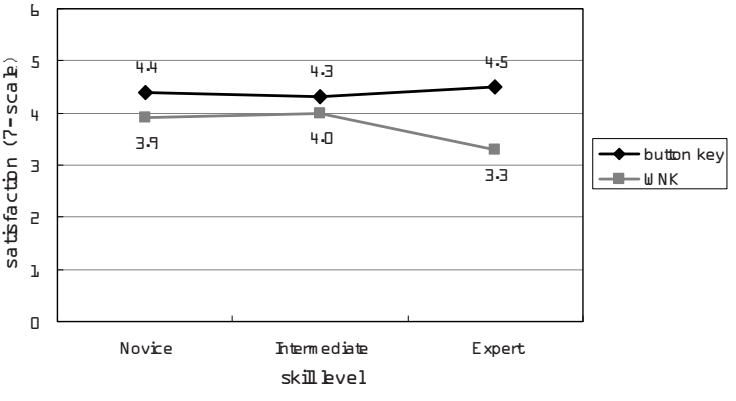


Fig. 10. Subjective rating for satisfaction for mp3

4 Discussion

Regarding the speed of performance, the wheel navigation key (WNK) helped subjects finish the task fast. This trend was more apparent among experts in both mobile phone and mp3. However, this does not always mean that WNK is faster than button key. The speed of performance varies depending on the combination of device type, skill level, and task difficulty. Therefore, the result was summarized and discussed as follows:

First, for mobile phone, the interface type did not affect the performance speed for easy task. The speed improvement was observed when WNK replaced the frequent button use for name search.

Second, for mp3, WNK helped subject finish the task quickly even for an easy task. It was because the mp3 always requires music selection which is a simple and repetitive searching task.

Third, expert performs better with WNK and prefer WNK to button key. It means that experts are be benefited from WNK in a simple menu selection task as well as long searching task.

Fourth, the task with four ways navigation had a low score in subjective rating. This result seemed to agree with the study by Kiljander (2004) who mentioned that the function of one-dimensional rotating wheel was easy to use but not two-dimensional navigation key.

Fifth, subjective rating score indicated that subjects showed a strong satisfaction when they successfully completed the task by using WNK. This could be a good sign for new product with WNK. However, it could be a natural response of users when they experienced unexpected benefit with a new device or interface.

In conclusion, the WNK is useful for expert and particularly for searching task with a large storage of information. However, since the device should cover various users with different skill levels, a designer should consider how to combine both keys to optimize the performance of mobile devices with complex and various functions.

References

1. Hix, D., Hartson, H.R.: Developing User Interfaces: Ensuring Usability Through Product & Process. John Wiley & Sons, New York, United States of America (1993)
2. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1994)
3. Kiljander, Juhani, H.: D.Sc.(Tech.).: Evolution and usability of mobile phone interaction styles. Dissertation University Teknillinen Korkeakoulu, Helsinki Finland (2004)
4. Ziefle, M.: The Influence of User Expertise And Phone Complexity On Performance, Ease of Use And Learnability of Different Mobile Phones. Behavior and Information Technology 21(5), 303–311 (2002)
5. Kim, Yong, J., Lee, Yeun, H., Choi, Cheol, Y.: The Structuring Process of Multi-Centered Usability Evaluation Method. Journal of the Ergonomics Society of Korea 24(2), 25–33 (2005)

Correlation Between Cognitive Style and Structure and Flow in Mobile Phone Interface: Comparing Performance and Preference of Korean and Dutch Users

Ji Hye Kim, Kun-Pyo Lee, and Im Kyeong You

HCIDL, Dept. of Industrial Design KAIST
373-1, Kusong-dong, Yusung-gu, Daejon, 305-701, South Korea
gimjee@gmail.com, kplee@kaist.ac.kr, cursa@naver.com

Abstract. This paper presents experiments conducted to determine the correlation between culturally different cognitive styles and issues of information architecture and flow, specifically in mobile phone interface. Korean and Dutch participants participated in on-screen prototype test and cognitive style test. In Experiment 1, each cultural group showed a different preference on the function/theme-related menus and individuals' categorization styles had correlation with their preferences. Overall, the findings indicated that performance and preferences in a certain menu structure are associated with cognitive styles and it eventually helps to design culturally adapted interface. In Experiment 2, both groups showed more favorable attitude toward a Parallel approach and no significant correlation between cognitive styles and performance or preference were found. The correlation between prior experience and preference was not found to be significant in any tests.

Keywords: User interface Design, Mobile Phone, Cognitive Style, Cultural Difference.

1 Introduction

Interest in the influence of culture on user interface design has been growing as the world market is globalized. However, products and services have been localized at the superficial level through checking elements like text, symbol and functionalities without reflecting any unconscious cultural effects. Moreover, interface studies with cognitive and implicit viewpoints are rare, so mobile phones seem to keep consistent interfaces relatively independent of culture.

Most of studies on cultural interfaces are based on cultural dimensions suggested by Hall(1981) and Hofstede(1980). However the frameworks seem to be outdated and still ambiguous since those are focusing on general values or norms in human society.

It is, therefore, essential that we understand the cultural impact on the mobile interface independently and examine how the mobile interface can be culturally adapted. This study limits the field of cultural differences to culturally influenced cognitive differences and aims to illustrate how culturally different cognitive styles

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 531–540, 2007.
© Springer-Verlag Berlin Heidelberg 2007

influence information structure in the mobile phone interface by examining users' performance and attitude toward the interface.

2 Theoretical Background

Anthropological and psychological studies continue to suggest that cognitive style is culturally different. Richard Nisbett(2004)'s studies report actual proves of such cultural difference. A central idea to his research is 'Holistic versus Analytic thought'. Holistic thought engages in context-dependent and holistic perceptual processes by attending to the relationship between a focal object and the field. On the other hand, Analytic thought engages in context-independent and analytic processes by focusing on a salient object independently from the context in which it is embedded.

Richard Nisbett and his colleagues have found cognitive differences between East Asian and Westerners. According to his investigation, East Asians attend more to the field and the relationship between an object and the field than Westerners (Masuda, Nisbett, 2001) and East Asians are inclined to explain events with reference to interaction between the object and the field, while Westerners are more inclined to explain the same events with reference to properties of the object. About ways of organizing the world, East Asians tend to group objects on the basis of similarities and relationship among the objects whereas Westerners tend to group the objects on the basis of categories and rules(Chiu, 1972; Unsworth, 2005). These researches indicate that thinking styles of East Asians are different from those of Westerners. In other word, East Asians relatively engage in holistic thoughts while Westerners engage in more analytic thought and approach.

A number of studies linking the cognitive distinctions with (hypermedia) interface styles have been carried out. Rau(2004) examined the cultural difference in the computer performance of Chinese and American users. The results indicated that a concrete representation with thematic structure was advantageous to the Chinese users, who are known to be more context-dependent and classify things on the basis of their relationships. Ford and Chen(2000) examined the effect of cognitive styles on hypermedia learning and found significant differences in navigation strategies used by Field-independent and Field-dependent learners. They found that Field-independent learners use the more analytical approach and Field-dependent learners use the more holistic approach.

3 Hypothesis

As the related studies indicate, we predicted that users' performance and favorable attitude would be enhanced when an interface is compatible with a cognitive style (in matched condition). To grasp which cognitive style has a correlation with which element in mobile phone interface, interface elements of mobile phone were conceptually divided into three different layers; Representation, Menu structure and Interaction flow. And then, the corresponding cognitive process in each interface layer and analytic/holistic characteristics in the cognitive process were listed as follows.

Table 1. Interface layers and related cognitive styles

Interface Layer	Cognitive process	Analytic	Holistic
Representation (component and template)	Perception	Field independent Verbal	Field dependent Visual
Menu structure	Categorization	Taxonomic Inferential Rule-based	Relational Contextual Family resemblance
Interaction flow	Task handling	Planned Organized Sequential Linear	Flexible Spontaneous Random Parallel

In this paper, we focused on 'Menu structure' and 'Interaction flow' layers and the cognitive styles related to the layers. So to speak, we hypothesized that taxonomic or relational tendency in categorization would have a correlation with a type of menu structure in mobile phone interface and linear or parallel tendency in task handling would have a correlation with a type of interaction flow in mobile phone interface.

3.1 Menu Structure and Categorization Style

An option(menu) for setting certain content in a mobile phone is classified as a 'setting' because of its function and also contextually attached to a menu containing the content itself because it is one of operations conducted with the content. For instance, to perform a task of setting(changing) a ringtone , you may go through a 'setting' menu or a 'sound' menu in the main screen. A 'setting' menu is functionally grouped by the common function of setting, whereas a 'sound' menu is thematically grouped by the shared context(or theme) of sound. Considering cultural cognitive difference, we predicted that they would show different performance and attitude toward certain menu structures. Specifically,

Hypothesis 1. East Asians would associate the task of 'setting content' more with a thematically grouped menu than a functionally grouped menu and show more favorable attitude toward the approach. On the other hand, Westerners would associate the task more with a functionally grouped menu and prefer the approach.

3.2 Interaction Flow and Task Handling Style

Mobile phones have a limitation on displaying information in parallel due to small-sized screens so it is important to properly organize and display the information in a page. Especially, for tasks constructed with a series of actions like sending SMS, the information flow should be carefully considered. Considering the cognitive difference, we predicted that they would show different performance and attitude toward certain interaction flows because of such cognitive difference. Specifically,

Hypothesis 2. East Asians would show enhanced performance and attitude in parallel flow than Westerners, while Westerners would have performance and preference enhanced in sequential flow than East Asians.

4 Experiment

For comparative experiment, Korea and the Netherlands were chosen for representative Eastern and Western cultures. This selection meets Nisbett's division of Eastern and Western cultures. Moreover, according to Hostede(1980)'s Index on national cultural difference, Korea and the Netherlands were proved to be relatively different in terms of IDV(Individualism vs. Collectivism index) and PDI(Power distance Index), which means that they belong to quite different cultural areas.

4.1 Procedure

Cultural Background Questionnaire was conducted to collect demographic data and eventually extract participants with relatively strong cultural identity. It aimed to collect 30 available data from each country by controlling external variables such as age and educational background. And then, Experiment 1 for verifying the hypothesis 1 was followed by Experiment 2 for the hypothesis 2. The entire experiment took 30 min. approximately.

All materials in the experiment were made in Korean language for Korean participants and English for Dutch participants. Korean participants were mostly students at KAIST and Dutch participants were students or faculty staffs at TU Eindhoven. They were randomly involved in the experiment through WWW.

4.2 Experiment 1

Prototype Test. A mobile phone prototype is an interactive prototype running in desktop environment which allows participants to perform tasks by simple mouse-clicking. It was built with basic interface elements which are closely related to the research questions.

In the mobile phone prototype, the main screen was consisted of six menus (Call history, Messaging, Phonebook, Sound, Display, and Settings) and setting content was possible through not only thematic menus like 'Sound' and 'Display' and but also a 'Setting' menu.

Participants were asked to change the ringtone(Task 1) and wallpaper(Task 2). After completing the two tasks, they were asked to perform the same tasks again in the other way(4 tasks in total). It was to let the subjects experience and compare two different ways so that a preferred approach could be asked at the end of the test.

Cognitive Style Test. Cognitive Style Test is to find out whether an individual cognitive style is taxonomic or relational and a methodology similar to Unsworth(2005)'s experiments were used. 26 sets were selected which were judged to have relatively high image-quality and clarity. One target picture and two alternative pictures were presented together and participants were asked to pick one alternative which goes the best with the given target picture as quickly as possible. The two alternatives were composed of one belonging to the same taxonomy as the target picture and the other one sharing a relationship with the target picture.



Fig. 1. Two approaches of setting wallpaper (Top: 'Setting' menu, Bottom: 'Display' menu)

Result. The collected data were analyzed by using SPSS and the significance level was $p<.05$.

Menus which they selected. Fisher's exact test was conducted with a 2 x 2 crosstable of cultural factor (Korean, Dutch) and selected menu (Setting/Sound, Display). As shown in Fig. 2, over 70% of both cultural groups set ringtone by using a 'Sound' menu and the results indicated that there were no significant difference in selected menus between the two groups($p=1.00$). For the task of setting wallpaper, most of the participants changed wallpaper by using a 'Display' menu and no significant differences were found between the groups($p=.52$). As a result, there were no cultural differences in menus they started with to perform the tasks.

Menus which they preferred. 53% of Dutch participants (n=16) preferred the 'Setting' menu and 77% of Korean participants (n=23) preferred the 'Sound' menu. Preferred menus in changing the ringtone were different between the groups ($p=.03$). In changing wallpaper, 53% of the Dutch participants (n=16) preferred 'Setting' menu and 73% of the Korean participants (n=22) preferred the 'Display' menu. The cultural difference did not appear statistically significant ($p=.06$) but, it was almost the same tendencies as national tendencies found in the former task.

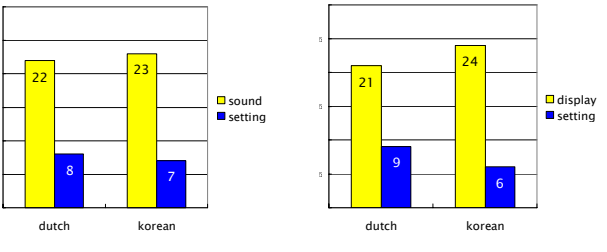


Fig. 2. Selected menus in setting ringtone(Left) and wallpaper(Right)

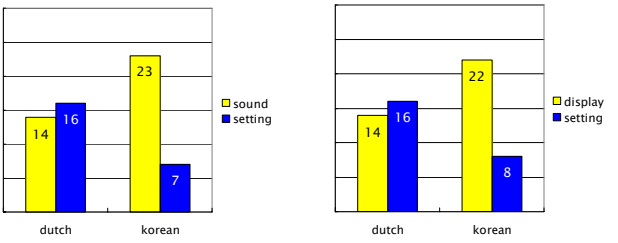


Fig. 3. Preferred menus in setting ringtone(left) and wallpaper(right)

Categorization style of each cultural group. Data taking longer than mean value of completion time(2.39s) were excluded since they were not thought to be unconscious and instant responses. After that, the percentage of each response type (relational grouping or taxonomical grouping) was counted so that an individual categorization tendency was yielded in a relative index ([100]: strong taxonomic ~ [1]: strong relational tendency).

To see if there are any differences in the categorization tendency between cultural groups or selected menus, 2-way ANOVA was conducted. An interaction effect between national groups and selected menus were not found. Korea group (M=35.73, SD=26.45) had more relational tendency than Dutch group (M=42.32, SD=30,70) but, the difference was not significant [F(1,56)=.74, p=.39].

Correlation between categorization style and selected/ preferred menu. However, the categorization tendency was found to be different by selected menus. The categorization tendency was different [F(1,56)=5.05, p=.03] between a group selecting the 'Setting' menu (M=53.35, SD=34.46) and a group selecting the 'Sound' menu(M=34.26, SD=25.02) in changing ringtone. Such tendentious differences were also found in changing wallpaper [F(1,56)=7.87, p=.01] between a group selecting 'Setting' menu(M=57.59, SD=31.50) and a group selecting 'Display' menu (M=32.84, SD=24.99). Categorization tendency was also found to be different by preferred menus. Categorization tendency were different [F(1,56)=16.86, p=.00] between a group preferring the 'Setting' menu in changing ringtone(M=56.78, SD=32.90) and a group preferring the 'Sound' menu(M=27.99, SD=18.78). Such tendentious differences were also found in changing wallpaper [F(1,56)=9.27, p=.00] between a group preferring the 'Setting' menu(M=53.01, SD=34.46) and a group preferring the 'Display' menu (M=29.70, SD=19.38). So to speak, the group which selected/preferred the 'Setting' menu had a tendency to be more taxonomic than the group which selected/preferred the 'Sound' or 'Display' menu in the both tasks. Then, if we know individuals' categorization tendency, can we find out menus that the individuals will select/prefer? A logistic regression analysis was conducted to reveal the casual relation.

A group for the 'Setting' menu was coded as 0 and groups for the 'Sound' or 'Display' menu were coded. The classification accuracy of a regression model with selected menu in changing ringtone and categorization tendency was 76.7% [chi-square=4.92, df=1, p=.03] and 76.7% [chi-square=8.32, df=1, p=.04] for selected menu in changing wallpaper and categorization tendency. In Table 2, B values were

shown to be (-). This means, the higher taxonomic person s/he is, the more possibility to select a 'Setting' menu in changing ringtone or wallpaper. Consequently, the higher taxonomic person s/he is, the more possibility to prefer a 'Setting' menu in changing a ringtone or wallpaper.

Table 2. Result of logistic regression analysis

	B	B	B	B	B	B
Cognitive style -> Selected menu : Ringtone	-.023	-.023	-.023	-.023	-.023	-.023
Cognitive style -> Selected menu: Wallpaper	-.030	-.030	-.030	-.030	-.030	-.030
Cognitive style -> Preferred menu: Ringtone	-.040	-.040	-.040	-.040	-.040	-.040
Cognitive style -> Preferred menu: wallpaper	-.031	-.031	-.031	-.031	-.031	-.031

Impact of prior experience. As a result of Fisher's exact test, no correlation was found between the ways their current mobile phones offer for ringtone setting and their selected menu during the test (p=1.00). There was also no difference between the ways of setting wallpaper in their current mobile phones and the ways they performed in the test (p=.73).

4.3 Experiment 2

Prototype Test. Two approaches (Linear and Parallel) were built for the 'Message' and 'Phone book' menus. A linear approach was a step-by-step process. On the other hand, in a parallel approach, all the items were shown in one scrollable page(Fig. 4).

The participants were asked to send SMS (Task 1) and save a contact (Task 2) in two approaches (4 tasks in total) and the two approaches were presented in different order for the two tasks to minimize a learning effect. Task completion time and a number of mouse clicks were measured and a preferred approach was asked after completing the 4 tasks.

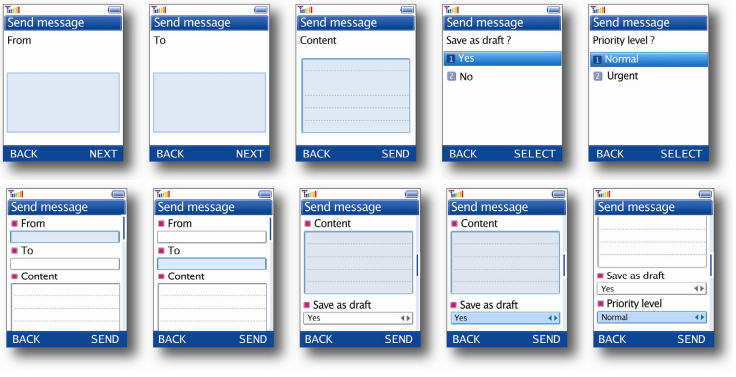


Fig. 4. Two approaches of sending SMS (Top: Linear, Bottom: Parallel)

Cognitive Style Test. To examine their cognitive styles which are especially related to task handling, a questionnaire with 16 statements was made by extracting statements on a Sequential-Global distinction from Felder's Learning Style Index and statements from Kaufman(1999)'s Polychronic Attitude Index and converting them into 5 point Likert scale.

Result. The collected data were analyzed by using SPSS and the significance level was $p<.05$.

Completion time. 2-way ANOVA showed that no interaction effect between cultural groups and approach types (T1: [$F(1,116)=.35$, $p=.56$], T2: [$F(1,116)=1.52$, $p=.22$]). It means that two cultural groups showed the same tendency across the two approaches. Korean participants performed faster in all tasks than did Dutch participants.

Mouse clicks. Two cultural groups showed the same tendency in the number of mouse clicks (T1: [$F(1,116)=.01$, $p=.92$], T2: [$F(1,116)=1.41$, $p=.24$]). No significant differences were found not only between groups but also between interface types.

Preference. The result of Fisher's exact test (two cultural groups × two approaches) showed that there was no difference in preference between groups (T1: $p=1.00$, T2=0.60), as Figure 7 shows. Interestingly, over 60% of each group preferred the parallel approach.

Cognitive style of each cultural group. A factor analysis yielded two factors; Random-Sequential factor and Multi Tasking–Single Tasking factor (excluding 5 statements, Cronbach α = 0.755, 0.725 respectively). The Korean participants averagely got higher values than the Dutch participants in the Random-Sequential factor [$F(1,58)=8.84$, $p=.004$] and the Multi Tasking–Single Tasking factor [$F(1,58)=12.52$, $p=.001$]. In other words, the Korean participants had more sequential and single-tasking tendencies in task handling.

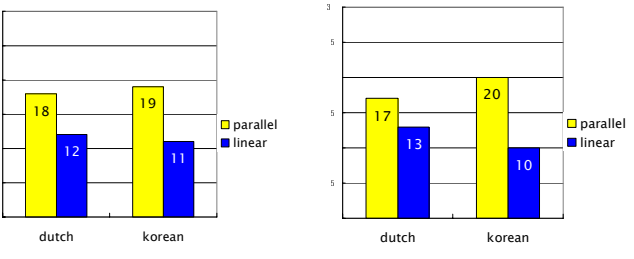


Fig. 5. Preferred approaches in sending SMS(left) and saving contact(right)

Correlation between cognitive style and performance/ preference. The correlation between such individual styles and individual performances(completion time, the number of mouse clicks) was not significantly found. Between preferred menus in the task of sending SMS(e.g., between a group preferring the linear approach and a group

preferring the parallel approach), there was significant difference in the Random-Sequential factor [F(1,58)=4.28, p=.04], but no difference was found in the task of saving contact. In the Multi-Single factor, no significant difference was revealed. Consequently, this finding indicated that performance and preference in the two different approaches(Linear, Parallel) hardly had a connection with cognitive styles on task handling(Random-Sequential, Multi-Single).

Impact of prior experience. The effect of prior experience was checked in the same way as used in Experiment 1. The ways of sending SMS and saving contact in the mobile phones participants currently use did not have any correlation with the approaches they preferred in sending SMS and saving contact(T1: p=.52, T2=.24).

5 Discussion and Implication

In Experiment 1, cognitive styles in categorization(Relational or taxonomic) had a significant correlation with types of menu structure(Thematic or Functional). Relational-grouping participants(Korean) were more likely to select and prefer the thematically grouped menu, whereas taxonomic-grouping participants(Dutch) tended to select and prefer the functionally grouped menu. Correlation between prior experience and performance & preference were not found significant so that the association between cognitive style and menu structure got more supported.

In Experiment 2, no correlation between cognitive styles in task handling(Sequential or Random, Multi Tasking of Single Tasking) and types of interaction flow(Linear or Parallel) was found. We speculated that the tasks(sending SMS, saving contact) were too basic and familiar so that the types of approach could not have any influence on the performance. Considering that in other related studies they presented 'learning' tasks to participants and found some difference, the degree of difficulty in the tasks we presented was not appropriate to find explicit differences.

Moreover, the way of measuring cognitive styles might have had some limitations. The 5 point Likert scale statements were written in too general and direct manners, not situation-specific. So participants confuse what they want to be with what they really are.

6 Conclusion

This study shows a possibility of cognitively adapted interface by connecting cognition model and interface architecture. The participants having a taxonomic categorization style performed the tasks in the functional approach and preferred the approach in the situation where both approaches were available. Therefore, for Western users who are known to be more goal-oriented, it may be better to organize menus by goals or functions so that they feel certain of goal achievements. For example, we can organize main menus with functionally grouped menus such as 'Setting' and 'Download'.

However, there were limitations in the experiment. The experiment environment was desktop-based so it might not have been realistic enough to demonstrate mobile phone interface. And the subject groups were mostly students in their early twenties who are more easily adjusting themselves to change so it might be difficult to find

clear differences between cultures or individuals. Therefore, we will need to capture a big-enough sample with diverse generations to ensure the validity of data. Other issues apart from menu structure and flow in interface will need to be also considered and so will other possible products.

References

1. Chiu, L.H.: A Cross-Cultural Comparison of Cognitive Styles in Chinese and American Children. International Journal of Psychology 7, 235–242 (1972)
2. Ford, N., Chen, S.Y.: Individual Differences, Hypermedia Navigation and Learning: An Empirical Study. Journal of Educational Multimedia and Hypermedia 9(4), 281–312 (2000)
3. Ford, N., Chen, S.Y.: Matching/Mismatching Revisited: An Empirical Study of Learning and Teaching Style. British Journal of Educational Technology 32(1), 5–22 (2001)
4. Kaufman, S., et al.: The Polychronic Attitude Index: Refinement and Preliminary Consumer Marketplace Behavior Applications. American Marketing Association Winter Educators Conference Proc. Marketing Theory and Application 10, 151–157 (1999)
5. Masuda, T., Nisbett, R.E.: Attending Holistically vs. Analytically: Comparing the Context Sensibility of Japanese and American. Journal of Personality and Social Psychology 81, 922–934 (2001)
6. Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and System of Thought: Holistic vs. Analytic Cognition. Psychological Review, 108, 291–310 (2001)
7. Rau, P.L.P., et al.: A Cross Cultural Study on Knowledge Representation and Structure in Human Computer Interfaces. International Journal of Industrial Ergonomics 34, 117–129 (2004)
8. Unsworth, S.J., Sears, C.R.: Cultural Influence on Categorization Process. Journal of Cross-cultural Psychology 36(6), 662–688 (2005)

Incorporating JND into the Design of Mobile Device Display

Joo Hwan Lee[1,2], Won Yong Suh[2], Cheol Lee[2,*], Jang Hyeon Jo[2,3], and Myung Hwan Yun[2]

[1] POSDATA Co. Ltd., Seoul, Korea
[2] Department of Industrial Engineering, Seoul National University, Seoul, 151-744, Korea
[3] Samsung Electronics Co. Ltd., Seoul, Korea
iehis@snu.ac.kr

Abstract. The main purpose of this article is to incorporate the JND (Just Noticeable Difference) into the design of mobile device display, especially LCD display of mobile phone. JND is the difference threshold between stimuli that can be detectable by human sense. Thirty participants were employed for two experimentations in order to find out JND value of sensation for LCD. The critical design variables of LCD and the affective component of user satisfaction were investigated using AHP and regression analysis. Finally, the JND of design variables of LCD and its characteristics were investigated.

Keywords: Mobile device, Display, JND, LCD.

1 Introduction

As information technology has developed speedy and digital convergence has been accelerated, mobile device has used in many cases and its user has increased rapidly. The trend of mobile device development is smaller, lighter, and multi-functioned [1]. Especially a small size and low image quality of display penal are handicaps of usability and satisfaction. For example, since a mobile phone, one of the representative mobile device, has limited the physical user interface due to its small display penal and button in spite of many functions, many users have difficulties in using mobile phone [2]. Especially, the small display of mobile phone with high-resolution images could cause considerable users' dissatisfaction due to the unnecessary delay for calculating of unimportant scene feature below the threshold of human sense of sight [3]. In fact, all components of the display characteristics are not responsible for user satisfaction while user experience and the improvement of all components of display may not guarantee user satisfaction.

In case of the design of small display, the concept of JND (Just Noticeable Difference) might be useful in that it could suggest an acceptable range of design specifications of the display. It is expected that incorporation of JND into the design of mobile device display can reduce perceptual redundancy to make physical user interface of

* Corresponding author.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 541–549, 2007.
© Springer-Verlag Berlin Heidelberg 2007

the display more usable. However, the previous studies on incorporating JND into the design of mobile device display are relatively few while many studies are available in the area of the image processing, virtual reality, speech recognition etc. [4, 5, 6]. In this study, we aim to investigate critical design variables of display for user satisfaction using the concept of JND and to suggest the design guideline of the variables of mobile device display.

2 Method

2.1 Experiment 1: Evaluation of User Satisfaction

We explored 22 affective components from the previous studies on usability and evaluation of affective satisfaction of mobile phone [7, 8]. We also analyzed hierarchical structure of the extracted affective components utilizing AHP. AHP is a mathematical decision making technique that allows consideration of both qualitative and quantitative aspects of decisions [9]. Table 1 shows seven affective components that were screened by expert panel.

Table 1. Seven affective components

No.	Affective component	Definition
1	Brightness	the quality or state of being bright
2	Clearness	the degree of distinguishable between colors
3	Color sense	the image of color of display
4	Reality	the degree of reality of display
5	Cubic effect	the degree of cubic effect of display
6	Craftsmanship	the degree of craftsmanship of display
7	Satisfaction	the degree of user satisfaction

We also investigated various design variables affecting LCD image quality such as DPI, brightness, contrast, color, angle of vision, and a delay of response from related works [10, 11]. We analyzed hierarchical structure of extracted components utilizing AHP again. After expert review, table 2 presents finally extracted four design variables and their definitions.

Table 2. Design variables affecting LCD image quality

No.	LCD component	Definition
1	DPI	Dot Per Inch
2	Brightness	the quality or state of being bright
3	Contrast	the degree of difference in tone between the light and dark partials
4	Color(RGB)	the particular visual sensation produced in this way, depending upon the wavelength

We simulated experimental environment using a 17-inch LCD monitor of computer (Hewlett-Packard 1702, pixel 800*600) instead of a LCD of mobile phone. The experimentation was implemented in a dark room in order to avoid compounding effects resulted from uncontrollable factors. Thirty participants, who had normal eyesight and were twenties, were recruited from the graduate student population. Fifteen different images were prepared for the experiment. The sample image size was 3cm*4cm as same as the mobile phone LCD size. The outer area of the sample image was painted in black as shown in Figure 1 and the sample images were presented in random order. The participants were asked to execute the experiment away from 30cm simulating usage of mobile phones. Table 3 shows the definition and variation of each variable, and Table 4 shows the specifications of the sample images.

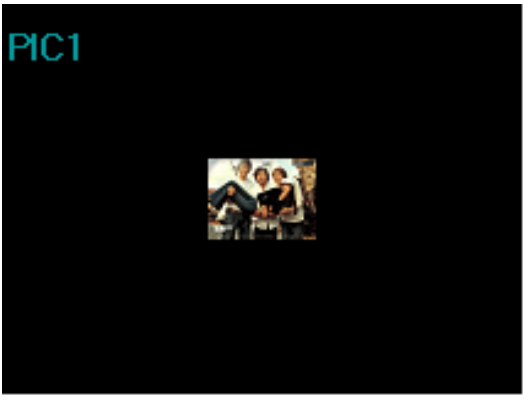


Fig. 1. The images were painted black outside the sample pictures

Table 3. The definitions and variations of each variable

Variables	The definitions and variations
DPI	DPI (dot per inch) is the number of pixel per inch and the index of resolution of display. From 36 to 200 DPI images were prepared for experiment.
Brightness	Brightness of images was measured by average of luminance.
Contrast	Contrast of images was measured by standard deviation of luminance.
Color (red, green, blue)	Color (red, green, blue) of images were prepared by variation of a quantity of red, green, blue.

2.2 Experiment 2: Measurement of JND

The concept of threshold is mentioned by the first stimulus strength to sensor. Threshold means the boundary of detectable sense. The minimum difference of two other stimuli is just noticeable difference (JND) [12]. Ernest Weber, a psychophysicist,

Table 4. The specifications of sample images

DPI	Brightness	Contrast	Red	Green	Blue
50	104.86	82.85	114.39	103.64	86.05
36	111.45	51.24	132.76	104.38	91.61
200	101.17	83.53	123.67	94.55	76.04
200	100.93	68.15	101.41	101.21	98.20
200	106.21	99.69	118.33	104.06	85.09
130	115.62	88.53	119.38	121.84	72.66
200	106.41	83.66	105.32	112.78	76.07
50	145.42	55.37	146.97	145.83	139.2
200	104.8	59.07	111.54	103.91	91.57
96	77.72	78.64	76.72	80.81	63.76
200	80.35	77.93	89.32	79.35	61.85
40	207.9	64.84	203.6	210.02	208.37
200	97.97	84.61	105.29	94.56	96.31
72	92.48	67.70	102.79	80.70	126.62
200	132.31	80.29	141.64	130.71	115.74

formulated $K=\Delta I/I$. K is Weber constant, ΔI is variation of the minimum difference of two other stimuli, and I is a reference of stimulus. Weber constant means that if a reference stimulus is larger, variance of stimulus is much larger. To the contrary, if a reference stimulus is smaller, variance of stimulus is much smaller.

The participants and experimental setup are prepared as same as experiment 1. We showed the images of 5 DPI and gradually higher by 20 DPI with an interval of one, and showed the images of 40 DPI and gradually higher by 150 with an interval of 10 for the sensitivity of DPI changes. We assumed the same Weber constant at other different DPI. We measured JND of DPI using mean value of three times measurement for each participant. Table 5 shows the scenario of experimental procedure.

We measured JND of 70.5 contrast (STD of luminance) and showed the images four times randomly to each participant as shown in Table 6. We measured JND of 92.54 brightness (average of luminance) and showed the images four times randomly to each participant as same as measurement of contrast. We prepared the controlled image of color (186 red, 171 green, 159 blue), and varied the color of images. The images were also presented randomly.

Table 5. The scenario of procedure for JND measurement of DPI

Image(DPI)	Ref.	off	6DPI	off	Ref.	off	7DPI	off	Ref.	off
Time(sec)	5	2	2	2	2	2	2	2	2	2
Image(DPI)	8DPI	off	Ref.	off	9DPI	off	Ref.	off	10DPI	off
Time(sec)	2	2	2	2	2	2	2	2	2	2

Table 6. The scenario of procedure for JND measurement of Contrast

Image(Contrast)	Ref.	off	74	off	Ref.	off	75	off	Ref.	off
Time(sec)	5	2	2	2	2	2	2	2	2	2
Image(Contrast)	76	off	Ref.	off	77	off	Ref.	off	78	off
Time(sec)	2	2	2	2	2	2	2	2	2	2
Image(Contrast)	Ref.	off	69	off	Ref.	off	68	off	Ref.	off
Time(sec)	5	2	2	2	2	2	2	2	2	2
Image(Contrast)	67	off	Ref.	off	66	off	Ref.	off	65	off
Time(sec)	2	2	2	2	2	2	2	2	2	2

3 Results

3.1 Experiment 1: Evaluation of User Satisfaction

The purpose of experiment 1 was to identify critical components to user satisfaction of LCD. We performed multiple regression analysis in order to investigate relationship between satisfaction of LCD and selected variables. Table 7 presents the result of multiple regression analysis. This model showed statistical significance ($p<0.0001$) with R-square = 0.84 and MSE = 23.58. Multi-collinearity is rather high due to color.

Table 7. The result of multiple regression analysis

Variable	Estimate Parameter	Standardized Estimated Parameter
Craftsmanship	5.53	0.44
Clearness	3.34	0.29
Reality	2.83	0.25
Cubic effect	3.41	0.31
Color sense	2.42	0.22
Brightness	-0.20	-0.40
Contrast	-0.21	-0.26
Color	0	0.12

We further analyzed partial test of each variable since some variables may have significant effect though result of multiple regression analysis indicated they were non-significant variables. Table 8 shows the partial model of DPI and contrast. The DPI and contrast model were statistically significant ($p<0.001$), and the interaction of other variables was not found at any significance level. The partial model of brightness was not statistically significant ($p=0.59$) and the interaction of other variables was significant at 0.1 significance level in brightness*contrast, brightness*red and brightness*blue. The partial model of color is statistically significant ($p<0.0001$) and the interaction of other variables was not significant at 0.1 significance level except for red*green. Table 9 shows that green, blue, green*blue, and red*green*blue were statistically significant factors.

Table 8. Partial model of DPI and contrast

DPI		Contrast	
Parameter	Estimated value	Parameter	Estimated value
Intercept	-12.45	Intercept	-67.12
DPI	1.41	lumin_std	3.05
DPI *DPI	-0.01	lumin_std*lumin_std	-0.02

Table 9. Partial model of color

Parameter	Estimated value
Intercept	1056.90
red	-3.24
green	-31.62
blue	14.62
red*green	0.00
red*blue	-0.09
green*blue	0.06
red*green*blue	0.00
red*red	0.06
green*green	0.14
blue*blue	-0.04

3.2 Experiment 2: Measurement of JND

The purpose of experiment 2 is to measure the JNDs of four variables identified by experiment 1. The measurement of JND is determined by 50% probability of change detection. The detection probability of each DPI, contrast, brightness and color (red, green, blue) were plotted and estimated. The estimated equation can calculate the point of 50% probability of detection. The calculated JND may inform of characteristics of each variable.

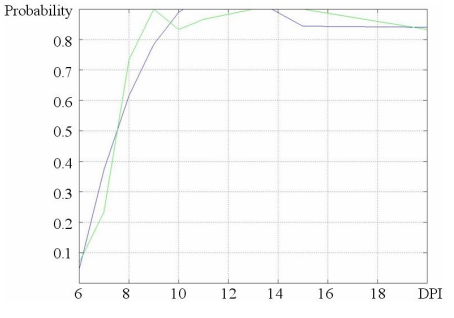


Fig. 2. JND of 5 DPI

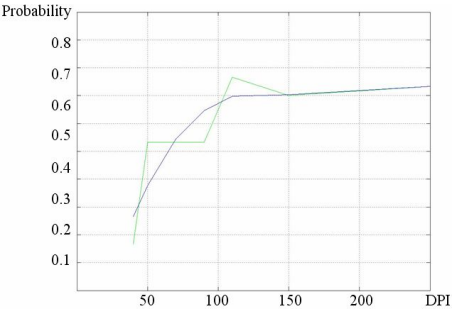


Fig. 3. JND of 40 DPI

Fig. 2 and Fig. 3 show the probability of DPI. The dotted line is the detection probability, and the solid line is the estimated probability. According to the result, the JND of 5 DPI is 6.94 DPI, and that of 40 DPI is 63.95 DPI. The Weber constant of 5 DPI is 0.39, and that of 40 DPI is 0.59. This result is not the same Weber constant. It is assumed that the relation between change of DPI and feelings of human is not multiple but rather quadric. If we know the DPI value that human cannot detect its change, we can calculate the acceptance range to measure JND of that value. Fig. 4 shows the probability of contrast. The JND of contrast is 64.20 in case of decrease, and 82.07 in case of increase. The Weber constant of decreasing contrast is 0.09, and that of increasing contrast is 0.16. If we adjusted less contrast in case of decrease than increase, it may be possible to reduce discordance of detection. Fig. 5 shows the probability of brightness. The JND of brightness is 80.26 in case of decrease and 106.08 in case of increase. The Weber constant of decreasing brightness is 0.13 and that of increasing brightness is 0.14.

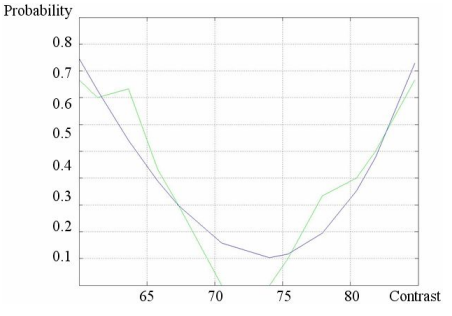


Fig. 4. JND of contrast

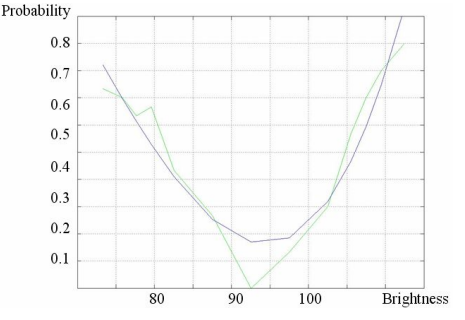


Fig. 5. JND of brightness

Fig. 6, 7, 8 shows the probability of color (red, green, blue) and Table 10 shows the JND of each color. Since color is composed of red, green and blue, the relative ratio of JND is significant for JND of color. Fig. 9 presents that JND of red is denser than green and blue. This means that human is more sensitive by change of red, and less sensitive by change of blue.

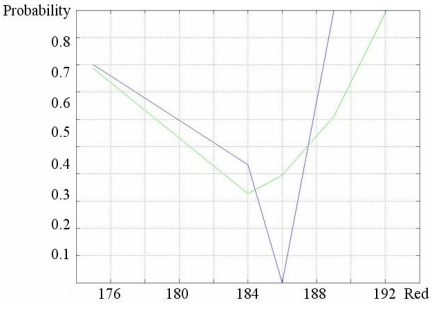


Fig. 6. JND of red

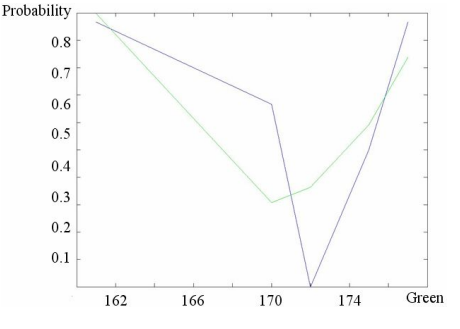


Fig. 7. JND of green

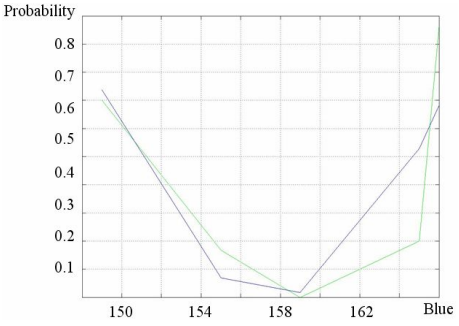


Fig. 8. JND of blue

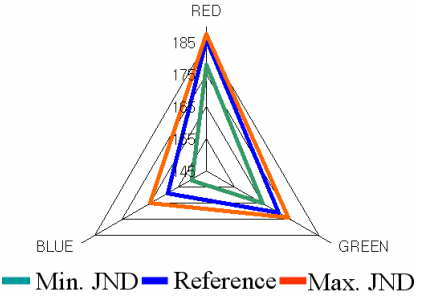


Fig. 9. Comparison of JND range of each color

Table 10. JND of color

Color	JND (decrease)	Reference	JND (increase)
Red	177.93	186.00	187.70
Green	165.01	171.00	174.04
Blue	150.53	159.00	164.80

4 Conclusion

The results of this study allow us to conclude that DPI, contrast, brightness and color (red, green and blue) are the critical design variables of the display of mobile phone for user satisfaction. The JNDs of the design variables were calculated and the characteristics of them were identified by the results of experiments. It is expected that the result of this study could be utilized to suggest design guidelines of mobile device design. There are many matters to be further investigated, such as experimental environment, types of images, individual difference, etc. In this study, we used a LCD monitor of computer instead of a LCD of mobile device so that potential discrepancy between related studies on JND might be found in result. Since the experiments in this study were implemented in laboratory environment, further study should be followed in various mobile environments considering inherent characteristics of mobile device.

Acknowledgements. This work was supported in part by the Research Institute of Engineering Science at Seoul National University.

References

1. Choi, Y.J.: The Evaluation Methodology of Mobile Phone Usability. Seoul National University (2005)
2. Jung, G.T., Che, I.S., Kwon, O.S., Lee, D.H., Kim, J.H.: Users Basic Characteristics for Designing the User Interface of Mobile Phone. IE Interface 15(1), 73–81 (2002)

3. Greenberg, D.P.: A framework for realistic image synthesis. Communications of the ACM 42(8), 44–53 (1999)
4. Ellis, S.R., Bréant, F., Menges, B.M., Jacoby, R.H., Adelstein, B.: Operator interaction with virtual objects: effects of system latency. In: Proc. 7th Int. Conf. on Human-Computer Interaction, pp. 973–976 (1997)
5. Mania, K., Adelstein, B.D., Ellis, S.R., Hill, M.I.: Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In: Proceedings of the 1st Symposium on Applied perception in graphics and visualization, pp. 39-47 (2004)
6. Buskirk, R.V., LaLomia, M.: The just noticeable difference of speech recognition accuracy. Conference companion on Human factors in computing systems, 95 (1995)
7. Han, S.H., Yun, M.H., Kim, K., Cho, S.J.: Development of a Usability Evaluation Method. Postech (1998)
8. Yun, M.H., Han, S.H., Hong, S.W., Kim, J.S.: Incorporating user satisfaction into the look-and-feel of mobile phone design. Ergonomics 46(13/14), 1423–1440 (2003)
9. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
10. Tombling, C., Tillin, M.: Innovations in LCD technology. Synthetic Metals 122, 209–214 (2001)
11. Menozzi, M., Lang, F., Napflin, U., Zeller, C., Krueger, H.: CRT versus LCD: effects of refresh rate, display technology and background luminance in visual performance. Displays 22, 79–85 (2001)
12. Goldstein, E.B.: Sensation and Perception. Wadsworth Publishing Company (1996)

Fit Evaluation of 3D Virtual Garment

Joohyun Lee[1], Yunja Nam[1,*], Ming Hai Cui[1], Kueng Mi Choi[2], and Young Lim Choi[1]

[1] Department of Clothing and Textiles, Seoul National University, Seoul, Korea
[2] Dept. of Fashion Design, Dong Seoul College, Seongnam-si, Korea
yunja@snu.ac.kr

Abstract. Fitting in the real world can be reflected in cyber space for 3D virtual fitting simulation technology to be used by the tool for fit estimating. This study examined objectiveness and correctness of the information on clothing fit using 3D virtual garment simulated. Subjects were selected in various BMI. The patterns (skirt & slacks) were developed in the sizes of each subject. The garments were constructed from the patterns. Sensory test was done to compare the virtual garment with the real garment and the vacant space between skin and garment was calculated. As a result, the appearance of the virtual garment did not express pulling and wrinkles. The vacant space of the virtual garment did not have influence upon the gravity to produce space at the place where the garment covered human body not to make actual appearance.

Keywords: fit evaluation, deviation, 3D virtual garment simulation.

1 Introduction

Consumers of today have strong desire for more various and differentiated products and service compared to the past. Based on analysis of individual customer and establishment of customer information, products can meet the desire of individual customer through mass customization. In the apparel and fashion industry, with the advancement of IT, the trend of customization is going on due to development of technology that encompasses 3D body measurement, virtual reality and CAD/CAM. Under these environments, 3D virtual garment system is now being developed with the worldwide attention to 3D technology. E-commerce is a new and exciting technology, attracting much interest. It has the potential of fundamentally changing the ways in which companies do business, thus having a profound effect on the management of the supply chain.

Some systems can create a 3D character that is similar to customer's body shape and then provide indirect experience by virtually simulated clothing selected by the customer and allowing the character to wear it as if she/he try it on by her/him self. The system, with combination of fashion industry with IT, is now utilized in e-commerce, virtual fashion show, and online fashion community on the Internet. 'My virtual model'(2006), in which the system is applied to Internet shopping mall, and

* Corresponding author.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 550–558, 2007.
© Springer-Verlag Berlin Heidelberg 2007

'Digital Fashion show'(2006), that provides solution related to virtual fashion show, and 'Stylezone' (2006), the online fashion community are among examples.

If these technologies can be utilized more efficiently in the production stage of garment, they could be used as a means to predict the final outcome of product even without manufacturing garment samples. Besides, we will be able to respond more effectively with the coming era of mass customization by making tailored clothing that considers each individual's body shape and taste. In addition, 3D virtual garment can be used to address the fit problem of clothing caused by the limited shopping environment.

The research related to 3D virtual garment system is now undertaking in several sectors, especially centering on development of technology linked to the development of software. In the field of Clothing and Textile, as a means to express clothing design, 3D virtual garment simulation is used. Some studies (Bae, 2004; Kim, 2000; Kim, 2002; Yoon, 2001) present a possibility that the system can be used in the planning step of garment product have been conducted.

However, in order to use the system as a tool for garment product planning or providing product information on e-commerce, the state of virtual garment of the system should be more accurate and real. Further, from the information provided by computer, human emotion can be predicted.

Thus, there is a need to verify the function provided by the system. In particular, the objectivity and accuracy of the fit information by pattern and body shape should be considered. Through this, studies to connect emotion to computer should be undertaken. So the purpose of this study is to review objectivity and accuracy of the fit information provided by the system through fit comparison in accordance with pattern and body shape of the virtual and the real garments. The specific studies are as follows: First, by using NARCIS, the 3D virtual garment system, functions and characteristics provided by the system can be figured out through 3D virtual garment simulation. Second, with regard to fit by pattern and body shape, objectivity for fit information provided by the system is reviewed by comparing the similarity of appearance between the virtual and the real garments. Third, objectivity and accuracy for fit information provided by the system can be reviewed by analyzing the deviation of the virtual garment from the real garment.

Through these review, this study is intended to provide basic data for developing technology to make 3D virtual garment system be used actively in clothing and textile research field in responding to the trend of mass customization and be utilized as a tool for planning of garment product and offering of product information.

2 Methods

2.1 Apparel Pattern Preparation

Based on BMI, subjects for the experiment were classified into three categories to reflect variety of body shape: Obese type, Normal type, and Lean type. The detailed information about subjects is shown in Table 1. Tight skirt and straight slacks are selected such items as experimental garments for the real and the virtual garments. The design of experimental garment was based on the simplest form of garment in

order to minimize errors caused by aesthetic factors and to operate virtual simulation. The apparel pattern of each item was manufactured on the basis of body size of each subject of the three body types. The apparel patterns were based on the prototype of Nam & Lee(1997).

Table 1. Subjects information

	Lean	Normal	Obese
Height	158.0cm	160.4cm	167.2cm
Weight	42kg	51.5kg	81.5kg
BMI	16.82	20.02	29.15
BMI=weight(kg)/height^2 (cm)*104 (Lean type∗19, 19′Normal type′24 Normal type, Obese type]24)			

2.2 Experimental Garments

By using NARCIS, six apparel patterns were simulated as virtual garment. NARCIS is a kind of software developed by D&M Technology in Korea and is used for virtual garment simulation by creating parametric body. It is used in this study due to its merit that can identify transformation of garment in case of wearing virtual garment. With NARCIS, parametric body was created in accordance with body shape and size of the three subjects. So all six patterns were simulated and a total of 18 virtual simulations were conducted.

2.3 Comparative Experimental Garments

For comparison, the real garment was made by using the same apparel pattern used in the virtual garment simulation. Materials used in making real garment are selected as materials that are similar to data for basic material provided by NARCIS.

2.4 Visual Analysis of the Virtual and the Real Garments

The visual evaluation of the virtual and the real garments were conducted through a sensory test. Panel is composed of twenty graduate school students who majored in Clothing and Textile: ten students for Ph.D. and the other ten for master' degree. For evaluation, the virtual and the real garments were presented as front, side, and back images. Participants replied as they saw the images appeared on the computer screen. The questions focused on whether there is difference between the virtual and the real garments or not, and each question was evaluated based on standard of Likert 5points.

2.5 Deviation Analysis of the Virtual and the Real Garments

To analyze the deviation of the virtual and the real garments, horizontal cross section of garments was created. The cross section of the real garment was created by overlapping 3D scan data that includes the state of wearing and not wearing a real garment. RapidForm2006 (Inus technology, Korea) was employed for creating cross section. The cross section of virtual garment was created by using coordinates value of dots of the cross section with the cross section image provided by NARCIS-DS.

AutoCAD2005 (AUTODESK, Inc.) was used to calculate vacant space distance and vacant space area between human body and garments using the cross sections of the virtual and the real garments.

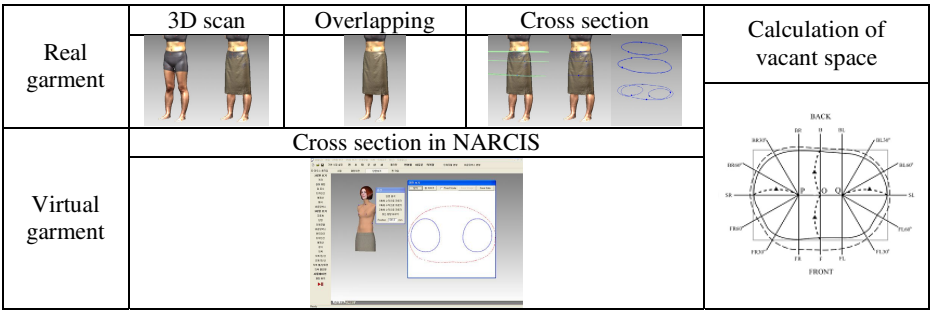


Fig. 1. Deviation analysis of the virtual and the real garments

3 Results

3.1 Virtual Garment Simulation

Parametric body was created and virtual garment simulation was conducted in NARCIS. The result is shown in Fig.2. The processes of creating parametric body and simulating virtual garment were relatively simple and speedy. Parametric body can be simply and quickly created as a body shape to represent many human bodies, and by changing major sizes of body, virtual human body model in various body shapes can be created. This virtual garment technology has a big possibility of utilizing in that it has an advantage to evaluate fit of garment product on-line.

In NARCIS, the simulated virtual garment was made just like the size of real garment for comparison. As a result, in the real garment, if the garment size was much larger or much smaller than the size of body, it is impossible to wear. But in the virtual garment, simulations for all 18 cases were possible. The result of the virtual garment showed that if the garment size was very big, the garment would not fit to the body and have much space to stay on air, while the garment was extended to be fitted if the size is much smaller. On the assumption that factors affecting these problems are regarded as strength operation of garment or physical property of garment itself caused by that, it can be figured out that strength and property of the virtual garment may differ from strength and property of the real garment. The 3D virtual garment system is necessary to evaluate garment fit objectively and accurately. But if this result was seen when it was used as a tool for evaluation, there would be a possibility to provide wrong information. So improvement is needed.

3.2 Visual Analysis of the Virtual and the Real Garments

The result from sensory test was analyzed by verifying significant difference in each question among 9 kinds of wearing cases using One-way ANOVA. As a result of

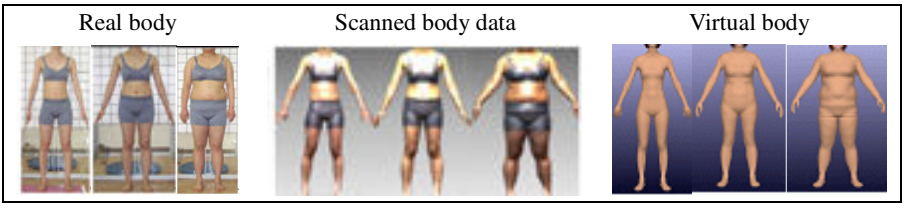
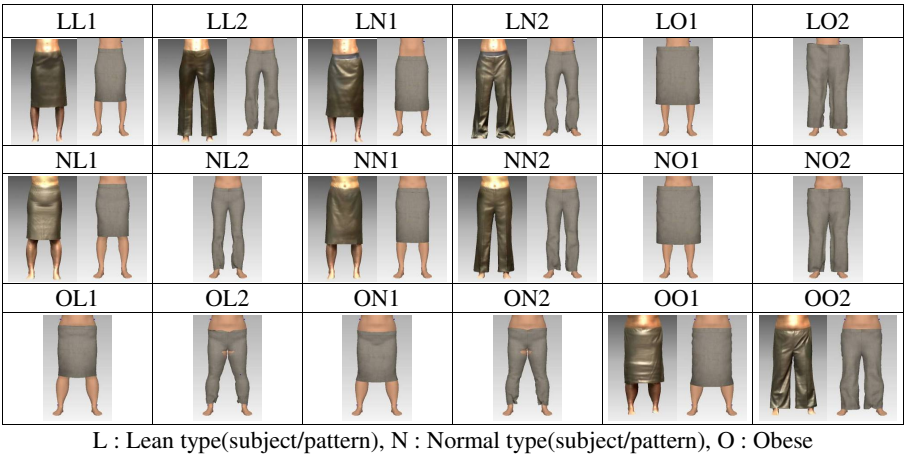


Fig. 2. Parametric body

verifying differences in 9 garments for each question, 16 questions showed difference at the level of 0.05. The result is shown as Table. 2.

Over all, in skirt, the virtual garment did not fit into the body. Because of this, the estimation of appearance in waist line didn't appear similar to that of the real garment. LN1 showed high score in the position of base line but low score in such items as pulling and wrinkles. So it confirmed that thick wrinkle created by space was not expressed well in the virtual simulation. OO1 recorded very low score in pulling and wrinkle compared to the other items. NLI was a case of having little space so the body surface shape was somewhat transformed by garment and showed severe pulling of textile caused by garment pressure. But expression on these parts was not shown in the appearance of the virtual garment so showed low score in the item related to vacant space area of the waist line and hip line. Compared with this, NN1 skirt showed high score in the items related with ease and position.

In case of slacks, LL2, NN2, OO2 showed low score in the question related to pulling and wrinkle while LN2 showed the lowest score in similarity of overall appearance and similarity of space of hip height line and thigh line, compared with other garment. In this case, space of slacks that includes relatively large space was shown in the shape of wrinkle and overall silhouette. Score related to these appeared to be very low so confirmed that its score in similarity of overall appearance was the lowest.



L : Lean type(subject/pattern), N : Normal type(subject/pattern), O : Obese type(subject/pattern), 1:skirt, 2:slacks
ex) LN1 is a case in which subject of lean type wear a skirt pattern for normal type.

Fig. 3. Comparison of the virtual garment and real garment

Table 2. Sensory test results of garment appearance

	LL1		LL2		LN1		LN2		NL1		NN1		NN2		OO1		OO2		F-Value
	M[a]	SD	M[a]	SD	M[a]	SD	M[a]	SD	M[a]	SD	M[a]	SD	M[a]	SD	M[a]	SD	M[a]	SD	
Ease amount of waist	**2.42**	1.01	**3.79**	1.18	**3.79**	0.91	**3.26**	1.32	**2.79**	1.31	**3.89**	0.99	**3.89**	0.99	**3.16**	1.01	**3.47**	1.12	4.028*
Ease amount of hip	**3.37**	1.21	**3.42**	1.17	**3.11**	0.93	**2.63**	0.95	**2.89**	1.24	**4.00**	0.88	**3.32**	1.25	**3.00**	1.29	**2.84**	1.01	2.515*
Ease amount of thigh	**3.32**	1.23	**3.21**	1.13	**3.05**	1.17	**2.32**	0.86	**3.32**	1.15	**3.68**	0.94	**3.26**	1.32	**2.79**	1.18	**3.16**	1.06	2.173*
Ease amount of knee	**3.05**	1.39	**2.95**	1.02	**3.11**	0.99	**2.53**	1.12	**4.00**	0.81	**3.79**	0.91	**3.47**	0.90	**2.47**	1.30	**3.47**	1.02	4.583*
Ease amount of hem line	**3.16**	1.16	**2.79**	1.18	**3.42**	1.01	**2.95**	1.02	**3.89**	1.04	**3.63**	1.01	**2.79**	0.97	**2.26**	1.14	**3.63**	0.83	4.684*
Position of waist	**3.89**	0.65	**3.79**	1.03	**3.47**	0.90	**3.11**	1.19	**2.63**	1.01	**3.11**	1.15	**3.26**	1.09	**3.05**	1.26	**3.42**	0.96	2.660*
Position of hip	**3.05**	1.07	**3.47**	0.77	**3.32**	0.94	**2.95**	0.97	**3.79**	1.08	**3.79**	0.91	**3.58**	0.96	**2.79**	0.91	**2.95**	0.97	2.974*
Position of hem line	**4.32**	0.82	**3.42**	1.12	**3.95**	0.84	**3.42**	1.12	**4.16**	0.83	**4.26**	0.56	**2.68**	0.88	**2.95**	1.12	**3.42**	1.07	7.213*
Position of seam line	**2.79**	1.09	**3.21**	1.27	**2.47**	0.84	**3.53**	1.07	**3.16**	1.06	**3.26**	1.19	**3.37**	1.01	**4.21**	0.63	**3.74**	0.73	4.628*
Length of dart	**3.16**	0.83	**3.74**	1.09	**3.16**	0.89	**3.84**	0.95	**3.84**	0.76	**3.68**	1.15	**3.95**	0.78	**2.58**	0.96	**3.84**	0.95	4.581*
Position of dart	**3.32**	1.00	**3.79**	0.97	**3.26**	1.14	**3.05**	1.17	**3.37**	1.01	**3.95**	0.78	**3.42**	1.07	**3.68**	0.94	**3.53**	1.07	1.420
Pulling and wrinkle from waist to hip	**3.21**	1.22	**2.42**	1.26	**2.11**	1.15	**2.47**	1.26	**2.00**	1.00	**3.11**	1.24	**2.74**	1.28	**1.79**	1.08	**2.68**	1.10	3.184*
Pulling and wrinkle from hip to knee	**3.21**	1.16	**2.42**	0.83	**2.53**	1.02	**2.32**	1.05	**2.32**	1.10	**3.16**	1.01	**2.32**	1.00	**1.84**	0.83	**2.95**	1.02	3.794*
Pulling and wrinkle from knee to hem line	**3.26**	1.04	**2.16**	1.06	**2.53**	1.02	**2.74**	1.28	**2.53**	1.38	**3.16**	0.95	**1.84**	0.95	**2.05**	1.02	**2.89**	0.99	3.845*
Silhouette front	**2.47**	1.07	**2.16**	1.25	**3.16**	1.01	**2.42**	1.17	**3.11**	1.28	**3.26**	0.93	**2.89**	1.04	**2.79**	0.97	**1.95**	0.70	3.612*
Silhouette side	**2.63**	1.11	**3.05**	1.07	**2.89**	1.10	**2.32**	1.05	**2.89**	0.93	**3.00**	1.00	**3.05**	1.17	**3.00**	0.94	**2.84**	1.06	0.369
Silhouette back	**2.63**	1.06	**2.42**	1.21	**3.05**	1.07	**2.26**	1.09	**2.89**	1.19	**3.32**	0.82	**2.58**	1.17	**2.00**	0.94	**2.74**	1.14	2.612*
Overall	**3.00**	0.88	**2.58**	1.01	**3.05**	1.02	**2.26**	0.80	**3.16**	1.11	**3.47**	0.84	**3.00**	0.88	**2.74**	0.73	**2.63**	1.95	2.977

[a]On a 5-point response scale with 1=Strongly Dissimilar, 2=Dissimilar, 3=Neutral, 4=Similar, 5=Strongly Similar.

* P≤0,05

3.3 Deviation Analysis of the Virtual and the Real Garments

In order to identify the deviation of the virtual garment from the real garment by pattern and body shape of the virtual garment, vacant space area and distance were calculated. As a result of analyzing vacant space area and distance, problems have been found that changes in surface of human body caused by garment pressure did not occur in the virtual garment and in case slacks contains lots of space, the phenomenon of pulling did not reflect. In other words, it can be confirmed that description of garment with out of size did not express well. The result is as following.

Table 3 and 4 show the result of vacant space area of the virtual and the real garments. Vacant space area of the real garment NL1 was calculated as negative value in waist part. In case of NL1, the body size was larger than that of garment caused the garment pressure in the body. This is because surface of body was pressed by garment pressure. However, in the virtual garment, pressing phenomena of body surface by garment was not expressed. In case of LN2, the vacant space area of the real garment appeared to be bigger.

Table 3. Vacant space area of the virtual and the real garments-Skirt

(mm^2)		Waist	Hip	Thigh	Mean
LL1	Real	1188	2671	10885	4915
	Virtual	3553	4514	13750	7272
LN1	Real	260	5600	14828	7678
	Virtual	2443	6879	16908	8744
NL1	Real	-2782	21	9350	2196
	Virtual	1999	2856	13705	6187

Table 3. (*Continued*)

NN1	Real	1797	4458	15217	7158
	Virtual	3788	6198	17838	9275
OO1	Real	3688	13639	19272	12200
	Virtual	3001	13193	24812	13668

Table 4. Vacant space area of the virtual and the real garments-Slacks

(mm^2)		Waist	Hip	Thigh	Knee	Mean
LL2	Real	872	2656	2716	5660	2976
	Virtual	1994	3034	3301	5353	3421
LN2	Real	2631	3365	9580	8477	6013
	Virtual	4266	5173	4195	5046	4670
NN2	Real	2406	1872	4242	7372	3973
	Virtual	2413	3927	5381	6663	4596
OO2	Real	6814	7824	4887	10075	7400
	Virtual	4680	15476	8924	13525	10651

Fig. 4 shows the difference between vacant space distances of the virtual and the real garments. That confirmed the phenomena in which surface of human body was transformed by garment pressure. This phenomenon was especially very well shown in NL1. In NL1, in case subject of normal type wore a skirt for lean type, negative value caused by garment pressure was shown in the real garment though effected by garment pressure was not expressed in the virtual garment.

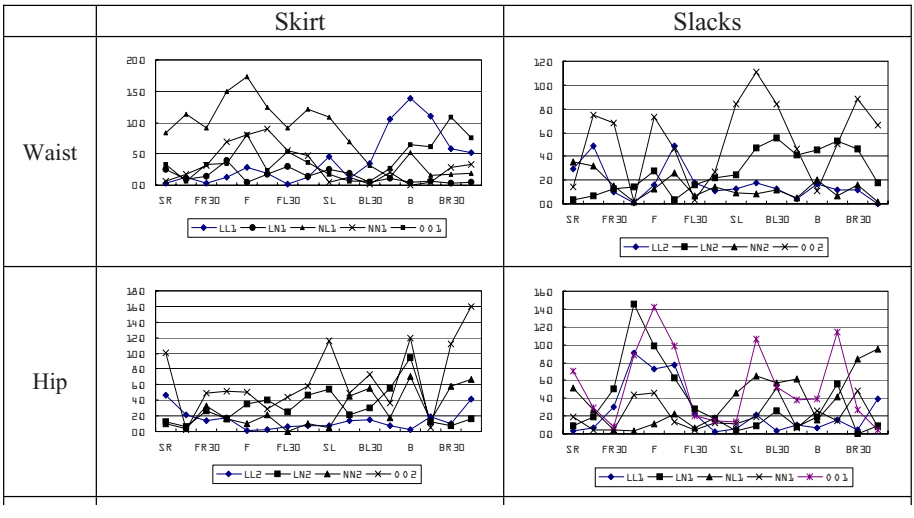


Fig 4. The difference between vacant space distances of the virtual and the real garments

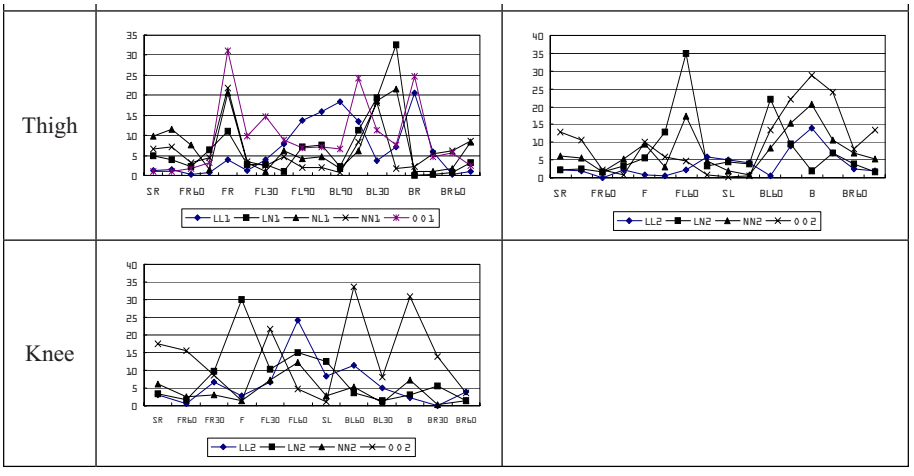


Fig 4. (*Continued*)

4 Conclusion

The study is intended to review objectivity and accuracy of fit information based on pattern and body shape in garment construction to verify functions provided by 3D virtual garment system. To this end, virtual body model is created by using 3D virtual garment system and then implement garment simulation. The study analyzed the result by comparing appearance and calculating the vacant space of the virtual garment with the real garment. The result is as following.

Using NARCIS, parametric body is created and then the virtual garment simulation is implemented. Based on this result, it is said that the process has a big possibility of utilizing as a tool to evaluate fit of garment product online in a relatively simple and quick way. But if it is not good to fit and impossible to wear in the real garment, simulation worked as if it is possible to wear in virtual simulation. This result may be caused by physics-based calculation and provide wrong information, so improvement is needed.

In appearance evaluation of the virtual garment, a simple skirt that has a few sewing lines showed high similarity in appearance and garment with adequate size has better appearance than garment with out of size. Particularly, garment with out of size had difficulty in anticipating fit through appearing because it does not describe well wrinkle and pulling caused by amount of space.

As a result of analyzing vacant space area and distance, problems have been found that changes in surface of human body caused by garment pressure did not occur in the virtual garment and in case slacks contains lots of space, the phenomenon of pulling did not reflect. In other words, it can be confirmed that description of garment with out of size did not show well.

Through garment appearance provided by 3D virtual garment system, both garment with adequate size and one with out of size should be well expressed in order to predict objective and accurate fit. In particular, to make a computer be a tool for

predicting human emotion, virtual outcome should be more objective and accurate than real one. Therefore, for development and use of technology related to 3D virtual garment system, work for reviewing utilization of various sectors needs to be done. And the result coming from this study will be basic materials for apparel and fashion field as useful and efficient data.

Acknowledgements. This work was supported by "Seoul R&BD Program" of South Korea.

References

1. Digital Fashion Show. December 06, 2006 from http://www.Dressingsim.com
2. Kim, H.Y.: A Study on the application of 3D digital animation model for fashion design 1. Journal of the Korean Society of Costume 50(2), 97–109 (2000)
3. Kim, J.E: Fashion design by 3D simulation based on characteristics and images of rainbow colors. Unpublished Ph. D. dissertation. Yonsei University (2002)
4. Bae, L.S.: A Study of Clothing Design in the Digital Age. The. Journal of the Korean society of costumes 54(4), 63–74 (2004)
5. Yoon, J.S.: Fashion Illustration thru the application of 3D computer animation technology. Unpublished master thesis. Ewha Womans University (2001)
6. Nam, Y.J., Lee, H.S.: Women's apparel pattern making. Kyohakyeongusa Seoul (1997)
7. Stylezone November 29, 2006 from http://www.Stylezone.com
8. Virtual Model. December 07, 2006 from http://www.mvm.com

Evaluation of Two Pointing Control Devices for a Cellular Phone

Ji Hyoun Lim[1], Cheol Lee[2,*], Sun Young Park[3], and Myung Hwan Yun[2]

[1] Department of Industrial & Operations Engineering, University of Michigan
1205 Beal Ave. Ann Arbor MI 48109-2117, USA
[2] Department of Industrial Engineering, Seoul National University, Seoul, Korea
[3] Korea Institute of Science & Technology Information, Seoul, Korea
iehis@snu.ac.kr

Abstract. Increasing number of functions in a cellular phone requires an advanced interface beyond a simple menu selection. In this study two pointing devices - an optic sensor and a pointing stick- were examined, which had recently been or planned to be applied to cellular phones. We evaluated two cellular phones equipped with each pointing device. Operations of the two devices were evaluated with objective and subjective measures. The throughputs, an index of performance based on Fitts' law, were collected in the multi-direction pointing and selecting task. A user interface (UI) checklist analysis was conducted as a subjective measure to evaluated users' acceptance of the devices. The results showed that values of throughput distinctively varied by direction of a movement when the optical sensor was used. In the UI checklist analysis, the pointing stick device was rated with higher scores than the optic sensor device.

Keywords: Fitts' Law, Pointing Control Device, User Interface, Mobile phone.

1 Introduction

A pointing control device has been essential in a graphic user interface (GUI) since Apple Macintosh introduced it in 1984 [1]. A GUI allows users to move the cursor to an item in a display, to select the item, and to activate the selected item by pointing and clicking. Various types of pointing devices have introduced to personal computing market following technological advances in designing and manufacturing control devices. Therefore, International Standard Organization (ISO) provides standard ergonomic requirements for office work with visual display terminals, especially for non-keyboard input devices (ISO9241-9) [2] to accommodate the user's biomechanical capability and to establish uniform guideline and testing procedures for evaluating computer pointing devices produced by different manufacturers [3]. Researchers have evaluated various types of pointing devices [1, 3, 4, 5, 6], including mouse, trackball, touchpad, and joystick using the concept of throughput, an index of performance based on Fitts' law [7]. However, the ISO9241-9 and the previous

* Corresponding author.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 559–565, 2007.
© Springer-Verlag Berlin Heidelberg 2007

studies were for the desktop or laptop computing environment. In this study we evaluated the two types of pointing devices equipped in cellular phones.

The two pointing devices tested in this study were an optic sensor controller and a stick controller. An optic sensor controls the cursor by position to position mapping. The optic sensor senses a position of finger which controls the system and maps the position of finger to the position of the cursor on the display. A stick controls cursor by force to velocity mapping. An example of this type of controller is the Trackpointe installed in IBM Thinkpad. The two pointing devices are installed in cellular phones both located at the center of main controllers (four buttons for left, right, up and down movements) which form circular shape and are placed between the display and the key-pad.

2 Performance Measurement

2.1 Throughput

Throughput is a rate of information transfer (in bits per second) when a user is operating an input device to control a cursor on a display. In the ISO9241-part 9, throughput is proposed as a metric to evaluate performance of pointing device for visual display terminals.

$$Throughput = \frac{ID_e}{MT} \quad . \tag{1}$$

$$ID_e = \log_2 \frac{d + w_e}{w_e} \quad . \tag{2}$$

Throughput is a function of index of difficulty (ID_e) and movement time (MT) for a given task. The index of difficulty is a function of movement distance (d) and target width of the displayed target (w_e).

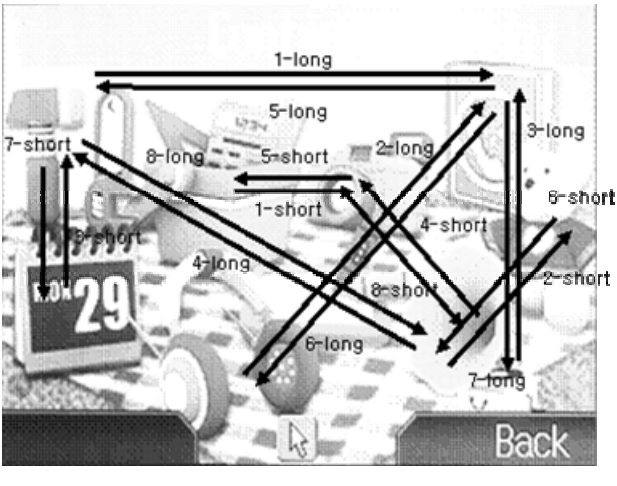


Fig. 1. The 16 types of pointing movements to measure throughput

To measure throughputs for the two pointing devices, movement times for 16 types of pointing movements (8 levels for direction and 2 levels for distance as shown in Figure 1) were collected from the 10 participants for each device (total number of trials = 2 devices * 8 directions * 2 distances * 2 trials = 64 trials/participant). To provide reality in use of the devices, the participants were asked to move the cursor displayed on the cellular phone screen by using the two pointing devices equipped in the cellular phones.

2.2 Results

Results showed statistically significant difference (p=0.013) in throughput by directions of movements. The types of devices and the movement distance alone didn't affect the amount of throughput. The result of repeated measure ANOVA is summarized in Table 1. The throughput of the optic sensor was 0.64 bits per second while the throughput of the pointing stick was 0.68 bits per second. Figure 2 showed varying throughputs depending on movement directions. The pointing stick showed similar amount of throughput across all directions. The optic sensor, however, showed noticeable differences between directions. The throughput in a movement toward left was higher than in a movement toward right. In addition, a horizontal movement shows 15% higher throughput than a vertical movement.

Table 1. ANOVA with repeated measures for Throughput (in bit/second)

Effect	*F*	*df*	*Sig.*
Device	2.296	19	0.15
Direction	4.185	13	0.013
Distance	0.344	19	0.56
Device * Direction	4.460	13	0.01
Device * Distance	153.916	19	<0.001
Direction * Distance	6.847	13	0.002
Device * Direction * Distance	5.585	13	0.004

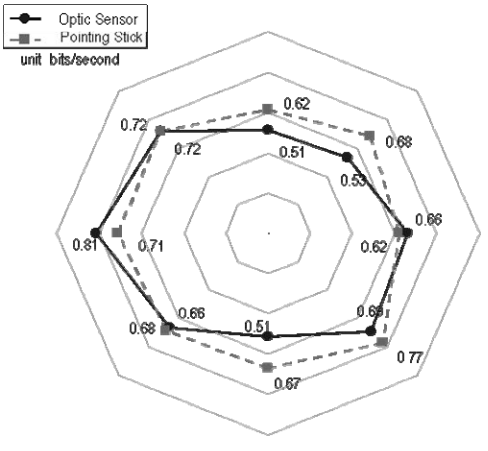


Fig. 2. Varying throughputs depending on movement directions throughput for the two devices

3 User Acceptance Measurement

3.1 User Interface (UI) Checklist Analysis

A physical user interface (UI) checklist was developed to assess users' acceptance of the two pointing control devices to activate functions of a cellular phone. The UI checklist has following three categories: performance in delivering a task, physical dimension, and demand in control. The performance category mainly focuses on the device's capability of carrying out requested tasks such as pointing and clicking. This category also includes items to evaluate feedback of an activated function. The physical dimension category is related to designing appearances of the control devices, such as configuration, shape, and size of the controller. The demand in control category examines the biomechanical and mental load resulted from using the control device. The structured items used to measure users' acceptance are presented in Table 2.

The UI checklist was rated by 10 graduate students majoring in human factors. The participants evaluated each item using a 7-point Likert-type scale (from 0 for not satisfied to 7 for fully satisfied) and specific descriptions supporting an evaluator's subjective rating of the item. Acceptance score was computed as a mean of the participants' ratings for all items in a given category of the UI checklist.

3.2 Results

The acceptance scores for the three categories were as follows: 4.9 and 5.2(optic sensor and pointing stick, respectively) for the performance category, 3.1 and 4.0 for the physical dimension category, and 5.6 and 6.4 for the interface category. The scores for the UI checklist are summarized in Table 2. The difference in acceptance of the two pointing devices for cellular phones was investigated with ANOVA. The result of ANOVA suggested that overall acceptance of a pointing device was significantly different (p<0.01).

Table 2. Summary of scores for the UI checklist

First level	Second level		Optic Sensor	Pointing Stick
Performance Category				
Pointing	Accuracy	End-point variation	3.7	3.8
		Gain* **	**2.6**	**4.6**
	Time	**Initiation time***	**4.0**	**5.1**
		Travel time	3.5	4.1
		Respond time	5.0	4.6
	Task-Completion Rate	Pointing error rate	4.1	4.6
Clicking	Accuracy	End-point variation	5.1	5.1
	Time	Activation time	5.0	4.6
	Task-Completion Rate	Clicking error rate	4.5	4.9
Feedback	Accuracy	Perceived link	5.3	5.7
	Time	Activation time	5.7	4.6

Table 2. (*Continued*)

Physical Dimension Category				
Configuration	Clicking	**Click separation*** **	**2.5**	**5.0**
		Click blocking*	**3.7**	**5.1**
	Harmony	**Seamlessness*** **	**3.1**	**5.3**
Shape	Outer Shape		5.0	5.8
	Overall Curvature		4.4	4.3
	Controller Curvature		4.0	4.4
Size	Outer (button) Size**		3.4	4.8
	Inner (sensor/controller) Size		5.0	5.4
Demand in Control				
Biomechanical load	Posture	Grasping	4.9	5.6
		Accessing*	**5.0**	**5.9**
	Operation	**Pointing actuation force***	**3.4**	**4.4**
		Clicking actuation force*	**3.9**	**5.0**
		Deviation and displacement	5.0	5.9
	Fatigue	Biomechanical fatigue	4.1	4.7
Mental load	Auditory Feedback	Control sound	5.0	5.0
		Signal sound	4.9	4.9
	Visual Feedback	Screen guidance	5.5	5.4
		Controller guidance	3.3	4.4
	Touch Feedback		4.0	4.9
	Fatigue	Pointing	3.9	4.3
		Clicking	3.3	3.8

Note: possible maximum score for each item was 7.0.
T-test: *p<0.05, **p<0.01, ***p<0.001

As shown in Table 2, across all the three categories rated, the pointing stick device was rated higher than the optic sensor device. The ANOVA result also suggested that the different categories showed significant different levels of acceptance on the two devices (p<0.001). For the performance category and the demand in control category showed relatively high level of acceptance for both devices than for the physical dimension category.

4 Discussion

In this study, two pointing devices were evaluated with objective – throughput, and subjective – UI checklist measures. Each method has its strength and weakness in evaluation of the emerging interface technologies for cellular phones. Although the objective measure required pre-determined tasks, large number of trial, and adequate number of participants, the results of analysis provided us a quantified performance index for the two pointing devices. The subjective measure, however, provided an expedite assessment [8] of user acceptance and highlighted the critical issues in adopting pointing devices in cellular phones.

As described, the throughput analysis showed that the values of throughput distinctively varied by direction of movement when the optical sensor was used. In the UI checklist analysis, the pointing stick device was scored higher than the optic sensor device. One possible reason why the optical sensor showed unbalanced throughput across movement directions was that the optical sensor requires more finger movements in operation. Unlike the pointing stick which senses force to the controller, the optical sensor senses positions of finger. Therefore, to make the cursor move further with the very limited size of optic sensor located at the middle of the main control buttons, users had to make multiple finger-movements.

Based on the quantified performance index of the two pointing devices, other designing factors could be examined. The limitation of optic sensor – requires multiple finger-movements, is possibly related to the low scores of the optic sensor device for pointing and clicking actuation force in the UI checklist analysis. Also, low scores of the optic sensor for click separation, click block, and seamlessness might attribute to the lower throughput of the optic sensor pointing device. Therefore, UI design should consider the objective performance measure of particular interface device so that the design of the control device could attenuate the weakness of technology or even select better technology for the control device.

One should be noted here is that the average throughputs of the two pointing devices equipped in cellular phones were much lower than other pointing devices for desktops or laptops. However, the average user acceptances of performance (rated in the UI checklist) were over 70% for both devices while the participants showed low acceptance of physical dimension (44% for the optic sensor and 57% for the pointing stick).

In summary, the throughput and the UI checklist analysis were practical measures to evaluate the two pointing devices in cellular phones. The pointing stick showed balanced throughputs across eight movement directions, and was rated with high scores from the UI checklist analysis.

Acknowledgments. This work was supported in part by the Research Institute of Engineering Science at Seoul National University.

References

1. MacKenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing device. In: Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI 2001, pp. 9–16. ACM, New York (2001)
2. ISO. ISO.DIS 9241-9: Ergonomic Requirements for Office Work with Visual Display Terminals, Non-Keyboard Input Device Requirements. International Organization for Standardization (2000)
3. 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 CHI ’99 Conference on Human Factors in Computing Systems, pp. 215–222. ACM, New York (1999)
4. Card, S.K., English, W.K., Burr, B.J.: Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT Ergonomics, vol. 21, pp. 601–613 (1978)

5. Epps, B. W.: Comparison of six cursor control devices based on Fitts' law models. In: Proceedings of the Human Factor Society 30th Annual Meeting. Santa Monica, CA: Human Factor Society, pp. 327–331 (1986)
6. Murata, A.: An experimental evaluation of mouse, joystick, joycard, lightpen, trackball, and touchscreen for pointing: Basic study on human interface design. In: Proceeding of the Fourth International Conference on Human-Computer Interaction, pp. 123–127. Elsevier, Amsterdam (1991)
7. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology: General 121(3), 262–269 (1992)
8. Ji, Y.G., Park, J.H., Lee, C., Yun, M.H.: A usability checklist for the usability evaluation of mobile phone user interface. International Journal of Human-Computer Interaction 20(3), 207–231 (2006)

Design and Evaluation of a Handled Trackball as a Robust Interface in Motion

Chiuhsiang Joe Lin, Chi-No Liu, and Jun-Lung Hwang

Department of Industrial Engineering, Chung-Yuan Christian University,
200, Chung Pei Rd., Chung- Li, Taiwan 32023, R.O.C.
hsiang@cycu.edu.tw

Abstract. In this study, a handled trackball was developed aiming at future use in a vibration environment within cockpits, ships, or other carriers. The study was to determine an optimal handle posture for the handled device from combinations of three forward slopes (0°, 15°, and 30°) and lateral slopes (0°, 15°, and 30°). The device was also compared with a table trackball for basic operation properties. An experimental cursor movement task was used to measure the response time of each design, accompanied by subjective fatigue and usability evaluations. The results found that the forward 30° and lateral 30° combination reached the top cursor movement performance without imposing undue fatigue to the operator. The study suggests using the forward 30° and lateral 30° handled trackball as the optimal design solution to maintain the performance when the operation of the trackball is under severe vibration environment.

Keywords: handled trackball, vibration environment.

1 Introduction

With the development of the information technology, computers have become an indispensable tool in our daily life and pointing devices, such as mice, trackballs, and touch screens, are particularly important to simplify the operation for users. Mice and trackballs are the most commonly used NKIDs (Non-Keyboard Input Devices) in the application of computer equipment. However, inappropriate operation or design of these devices may decrease performing efficiency and even bring about CTDs (Cumulative Trauma Disorders) to the muscles or bones of users after operation for a relatively long time, and the social, production, medical and human resource costs paid for CTDs are very high [4].

Most previous research assumed that computers were operated in a static environment with the mouse as the key control device for the operation. In their research on the operation of mouse, trackball, touch pad, tablet, and track point for linear motion, Accot and Zhai [1] found that the mouse required the least time for the linear motion, while it took the most time for the trackball to finish the motion. However, mouse operation is not suitable in every environment. It is not the best choice, for example, for radar or monitoring systems or other HCIs (Human Computer Interfaces) on ships. These systems operate in a non-horizontal or vibration

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 566–575, 2007.
© Springer-Verlag Berlin Heidelberg 2007

environment where users must balance themselves using additional support force [3] and, thus, have difficulties in operating the usual devices. It is therefore important to find out a way for stable operation of the HCIs. The trackball is usually a better choice than the mouse under motion environment, because it is fixed on the work surface and will not move with the external motion. However, when the body is subject to severe motion, there is a need for the hand to grasp on a firm object to maintain stability. At the moment of severe vibration, it is difficult to operate the table trackball. Accordingly, this study suggests using a handled trackball to provide support for users and facilitate smooth operation of the HCIs in a vibration environment. The trackball is placed on top of the handle and two buttons are positioned at the front of the handle as shown in Figure 1. The handled trackball is inserted in the armrest of the operator seat. The handled trackball is used in the same way as a normal trackball except that the handle provides secure grasp for the hand as the ship or carrier is subject to severe wavy movement. The idea is that the operator can be seated stably and the trackball can be operated with the thumb (for the trackball), index finger and middle finger (for the two buttons), while the ring and little fingers can grasp the handle to keep a stable posture of the hand and arm while the body is subject to motion.

How to design the angles and orientation of the handle becomes a critical design question in this study. A review of past studies on handled tool was first performed. Woodson et al. [10] found that the slope angle of the handle might affect the position of the wrist and forearm, making them fatigued or painful, so they suggested a forward slope of 0~10° to solve this problem. Some researchers suggested a proper grasp forward slope of 33° [5] and a handle with a minimum length of 100mm or an ideal length of 115~120mm [8]. In the research on the relationship between six diameters (25, 30, 35, 40, 45, 50mm) of handles and the grip strength, Kong and Lowe [7] found that users felt most comfortable and the highest grip strength value is acquired with diameters of 30, 35 and 40mm. A diameter between 38mm and 51mm resulted in the least muscle movement to operate a round handle [6]. It exerted higher force on the handle and could operate for a longer time before becoming fatigued. Brumfield and Champoux [2] further pointed out that a movement from 10° of flexion to 35° of extension was enough for the wrist to operate the handle in normal conditions, though it could move to a much larger extent.

Based on these recommendations and considerations, a trackball was instrumented on the handle with nine angle configurations. This study is designed to analyze the angle of the handled trackball. The study aims at finding a comfortable and productive position for the handled trackball.

2 Method

2.1 Participants

Six male and four female students (mean age: 25.3; mean height: 165.6cm) participated in the experiment. No participant was colorblind, suffered from eye diseases, or hand-arm problems. Each participant had a normal eyesight or at least 0.8 visual acuity after correction. The dominant hand was the right hand for all of them.

The participants were requested to fill in the "Personal Basic Information and Informed Consent Form of the Experiment" before the experiment commenced.

2.2 Equipment

This study utilized a desktop computer (Pentium □/1.70GHz, 256MB RAM) with a screen (Samsung 17" LCD, Model: SyncMaster 172B). The display resolution was set to 1024*768 pixel. The study used a table trackball (Macally, Model: Langend Ball) and a handled trackball (USB Geek, Model: Fish Handheld Mouse DH1), the latter was modified according to suggestions of ideal handle dimensions from the literature as the following: length 120mm [8], diameter 40mm [6], [7]. The handle angles were instrumented at three lateral slopes 0°, 15°, 30° and three forward slopes 0°, 15°, 30°. The experimental task was written in Microsoft Visual Basic 6.0 and executed in the Windows XP environment.

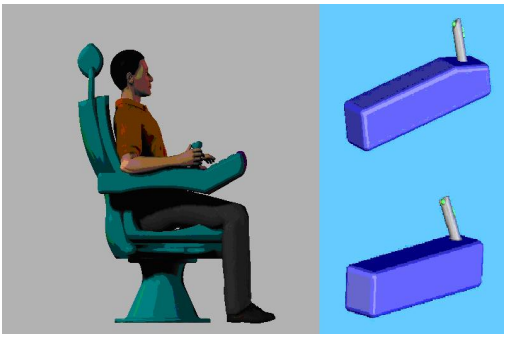


Fig. 1. The handled trackball inserted in the armrest of a seat, allowing the operator to stably operate the trackball under motion environment. Two handle slope configurations were shown here.

2.3 Experimental Task and Design

A cursor movement task was carried out with the experimental interface as shown in Fig. 2. A yellow question mark appeared in the center of the screen at the beginning of the experiment. A participant was requested to move the cursor using the pointing device to click the question mark button and activate the operation. A target shown as the red love heart symbol appeared randomly at one of the angles of 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315° and the participant was requested to move the cursor and click on the target as quickly as possible. The red heart symbol disappeared when it was clicked and the yellow question mark appeared again at the center of the screen. Then the participant continued the remaining actions as randomly appeared until the experiment was completed. The target appeared 5 times at each angle in a random order and a total of 40 (5 × 8) actions were to be completed for the experiment. The size of each red heart target was 0.8cm^2 and the distance between the target and the center was 11cm. The click time for each action was recorded by the system automatically.

There are three factors for the experiment, the target angle, forward slope angle, and lateral slope angle. The target angle contains the eight different angles, 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°. The forward and lateral slope each contains three angles 0°, 15°, and 30°. In the first part of analysis, the click time was analyzed against the three factors. For the purpose of comparison, the normal table trackball was added in the experiment. In the second part of analysis, the nine handle slope configurations and the table trackball were considered as one factor, namely the device type. The click time was analyzed against the target angle and the device type. The participant was treated as the block for the experiment. There were five replications due to the five occurrences at each target angle.

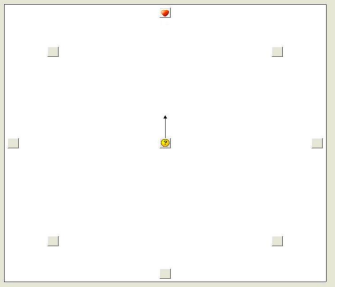


Fig. 2. The experimental task, with the question mark as the start and the red heart as the target, was to move the cursor as quickly as possible to the target and click on it

2.4 Subjective Evaluation

In addition to the task performance measure, the experiment result is enhanced with the subjective evaluation of the participants. An evaluation form is issued to each participant to complete after the experiment is completed. The participants are requested to rate the seven fatigue questions (neck and shoulder, right upper arm, right forearm, right wrist, right palm, right index finger, and right thumb) and four usability questions (hard to slide this trackball, hard to control this button, hard to grasp this sloped angle, and this performance is not good) on the form. The questions were rated with responses between one and five. A higher score indicates higher fatigue perception or higher dissatisfaction to the device usability.

3 Results

3.1 Results of the Experimental Task

In the first part of analysis of variance, the click time was analyzed against the target angle, forward slope, and the lateral slope angle. Both the target angle and forward slope were significant, but the lateral slope was not (Table 1). The click times at the target angles 0°, 90°, 180°, and 270° are shorter than those at the diagonals. For the three forward slopes, the shortest click time appeared at the 30° forward slope. Two interaction terms were significant, target angle by forward slope ($F(14,3519)=3.28$, $P<0.001$), and forward slope by lateral slope ($F(4,3519)=7.78$, $P<0.001$).

Knowing that the forward slope angle is the dominating factor, a combination of forward and lateral slopes must still be determined to reach a final design. In the second part of analysis, the click time was analyzed against the target angle and the device type which contains the nine slope combinations and the table track ball. The target angle and the device type were significant (Table 1). The interaction was also significant ($F(63,3911)=2.66$, $P<0.001$). The results of the target angle were similar to the first analysis. In this analysis, the shortest click time appeared when the handled trackball was operated at the forward-lateral slope combination of 30°-30°, while the longest click time appeared when it was operated at the forward-lateral slope combination of 0°-30° and 0°-0° (as confirmed by the Duncan test shown in Table 1). With this analysis, the final design slope angle was reached as the forward-lateral slope combination of 30°-30°. Compared to the performance of the table trackball, the click times for some slope configurations were higher and the others were lower. The click time values for the handle instrumented device were comparable with that of the table trackball, indicating that the handle instrumentation did not seem to change the pattern of use while providing additional hand grasp.

Table 1. Summary of Means, ANOVA, and Duncan test results for the two analyses

Click time by target angle and handle slope					Click time by target angle and device type of pointing device				
source	level	average	$F_{n,m}$	p-value	source	level	average	$F_{n,m}$	p-value
Target angle	270°	1.78^{A}	$F_{7,3159}$ 130.5	<0.001	Target angle	0°	1.79^{A}	$F_{7,3911}$ 106.6	<0.001
	0°	1.80^{A}				270°	1.82^{A}		
	90°	1.83^{A}				90°	1.86^{A}		
	180°	1.99^{B}				180°	2.01^{B}		
	315°	2.29^{C}				315°	2.27^{C}		
	135°	2.31^{C}				135°	2.33^{CD}		
	45°	2.44^{D}				45°	2.40^{D}		
	225°	2.63^{E}				225°	2.59^{E}		
Forward slope	30°	2.06^{A}	$F_{2,3519}$ 22.13	<0.001	Device type	30-30	1.99^{A}	$F_{9,3911}$ 7.46	<0.001
	15°	2.13^{B}				30-0	2.04^{AB}		
	0°	2.22^{C}				15-15	2.06^{ABC}		
						15-0	2.12^{BCD}		
						table	2.13^{BCD}		
Lateral slope	15°	2.10	$F_{2,3519}$ 1.09	0.338		30-0	2.14^{CD}		
	0°	2.12				0-15	2.14^{CD}		
	30°	2.13				15-30	2.20^{DE}		
						0-0	2.26^{E}		
						0-30	2.26^{E}		

* A, B, C, D, and E indicate the grouping by Duncan tests ($p<0.05$). The values with the same alphabet have no difference between their means.

3.2 Subjective Evaluation Results

Limited by the page allowance of the paper and the fact that the major interest of this study is to determine an optimal angle for the handled trackball as used in the setting described above, the results of the subjective evaluation were presented and summarized for the slope angle and the device type only.

3.3 Analysis of Fatigue Questionnaires

The analysis results of the fatigue questionnaires were summarized in Table 2. The forward slope had significant effects on the fatigue of the right upper arm, right forearm, right wrist, and right thumb. However, the lateral slope was not significant for any part of the body. The Duncan test (Table 3) further demonstrates that the forward slope of 30° brings about lower fatigue to the right upper arm, right forearm, right wrist, and right thumb, while the forward slope 0° brings about higher fatigue to these parts of the body.

When the slope angles were considered together, the device type had significant effects on the fatigue of the right upper arm, right wrist, and right palm (Table 2). The Duncan test result is presented in Table 4. The interest here is to check whether the

Table 2. Summary of the ANOVA results of fatigue questionnaires

Items	Forward slope	Lateral slope	Device type
Neck and shoulder	□	□	□
Right upper-arm	*	□	□
Right forearm	*	□	*
Right wrist	*	□	*
Right palm	□	□	*
Right index finger	□	□	□
Right thumb	*	□	□

*: P-value<0.05 □: P-value>0.05

Table 3. Duncan test results of forward slope on significant fatigue questions

Forward slope	Right upper-arm		Right forearm		Right wrist		Right thumb	
	Average	Duncan	Average	Duncan	Average	Duncan	Average	Duncan
30°	2.43	A	2.80	A	3.13	A	2.93	A
15°	2.83	AB	3.13	A	3.30	A	3.17	AB
0°	3.07	B	3.67	B	3.93	B	3.53	B

* A and B indicate the grouping by Duncan tests (p<0.05). The values with the same alphabet have no difference between their means. Lower values indicate less fatigue.

Table 4. Duncan test results of device type on significant fatigue questions

Device type	Right forearm		Device type	Right wrist		Device type	Right palm	
	Average	Duncan		Average	Duncan		Average	Duncan
15-15	2.50	A	15-15	2.70	A	15-0	2.30	A
30-30	2.60	A	30-0	3.10	AB	30-0	2.50	AB
30-15	2.90	AB	30-15	3.10	AB	30-15	2.60	AB
30-0	2.90	AB	30-30	3.20	AB	30-30	2.60	AB
15-0	3.10	ABC	15-0	3.40	ABC	15-15	2.70	ABC
0-15	3.50	BC	0-15	3.70	BC	0-30	2.80	ABC
0-30	3.60	BC	15-30	3.80	BC	15-30	3.10	ABC
15-30	3.80	C	0-30	3.90	BC	0-15	3.10	ABC
table	3.80	C	table	4.10	C	0-0	3.30	BC
0-0	3.90	C	0-0	4.20	C	table	3.50	C

* A,B and C indicate the grouping by Duncan tests (p<0.05). The values with the same alphabet have no difference between their means. Lower values indicate less fatigue.

best performance slope combination 30°-30° showed any adverse fatigue effect. It can be seen that the 30°-30° device was among the lowest fatigue group, while the table trackball and the 0°-0° device were among the highest fatigue group.

3.4 Analysis of Usability Questionnaires

The ANOVA results were summarized in Table 5. The forward slope had significant effects on three questions: hard to control this trackball, hard to grasp this sloped angle, and this performance is not good. The lateral angle was not significant in any of the questions. The Duncan test (Table 6) further demonstrates that the highest score appeared at the forward slope 0°. This indicates that the participants agreed that the handled trackball was not easy to operate or grasp when its forward slope was at 0°.

Table 5. Summary of ANOVA results of usability questionnaires

Items	Forward slope	Lateral slope	Device type
Hard to slide this trackball	□	□	□
Hard to control this button	*	□	*
Hard to grasp this sloped angle	*	□	*
This performance is not good	*	□	*

*: P-value<0.05 □: P-value>0.05

Table 6. Duncan test results of forward slope on significant usability questions

Forward slope	Hard to control this button		Hard to grasp this sloped angle		This performance is not good	
	Average	Duncan	Average	Duncan	Average	Duncan
30^0	1.90	A	2.77	A	3.00	A
15^0	1.97	A	3.07	AB	3.13	A
0^0	2.43	B	3.57	B	3.87	B

* A and B indicate the grouping by Duncan tests ($p<0.05$). The values with the same alphabet have no difference between their means. Lower values indicate lower unusability.

Table 7. Duncan test results of device type on significant usability questions

Device type	Hard to control this button		Device type	Hard to grasp this sloped angle		Device type	This performance is not good	
	Average	Duncan		Average	Duncan		Average	Duncan
30-30	1.80	A	15-15	2.40	A	15-15	2.70	A
15-30	1.90	A	30-0	2.60	A	30-30	2.80	AB
30-15	1.90	A	30-30	2.70	AB	15-0	3.00	AB
15-0	2.00	A	30-15	3.00	ABC	30-0	3.00	AB
15-15	2.00	A	15-0	3.10	ABC	30-15	3.20	ABC
30-0	2.00	A	0-15	3.10	ABC	0-15	3.70	BCD
0-15	2.30	AB	table	3.10	ABC	0-30	3.70	BCD
0-0	2.50	AB	0-30	3.70	BC	15-30	3.70	BCD
0-30	2.50	AB	15-30	3.70	BC	table	4.00	CD
table	3.00	B	0-0	3.90	C	0-0	4.20	D

* A,B,C and D indicate the grouping by Duncan tests ($p<0.05$). The values with the same alphabet have no difference between their means. Lower values indicate less fatigue.

The device type was significant on the usability questions, hard to control this trackball, hard to grasp this sloped angle, and this performance is not good (Table 5). The Duncan test (Table 7) further demonstrates that the 30°-30° device was among the lowest unusable group, while the table trackball and the 0°-0° device were among the highest unusable group.

4 Discussion

In the cursor movement experiment, the longest click time appeared at the forward-lateral slope combination of 0°-0° among all types of pointing devices. If we compared this with the result of the subjective evaluation, there seemed to be possible reasons for this poor performance. First, the fatigue questionnaires show that the highest rating for the right forearm and wrist at this slope configuration. This indicates that the participant had the strongest fatigue perception possibly resulting from the stress on the muscle of the right forearm and right wrist at this slope combination. Additionally, for the two questions "hard to grasp this sloped angle" and "this performance is not good", the ratings were, respectively, near 4 and beyond 4 points at this slope combination, indicative of uneasy grasp and poor performance. Based on careful observation of the grasping posture of the participant, it was found that when grasping the handled trackball in the posture and using the thumb to operate the device, the coupling between the palm and the surface of the handle was poor, bringing about stress and fatigue to the muscles of the palm and thumb. All of this confirmed that the 0°-0°slope resulted in the poor posture and therefore poor task performance.

It is interesting to note that the cursor task found that more time was required to click the target diagonally than to locate it horizontally or vertically, which is consistent with the finding of Thomas and Henry [9].

The study found that the forward slope was the dominating factor in determining a good grasp posture for the hand and arm. For the forward slope, the result of the shortest click time at 30° is similar to the argument of Hsu and Cheng [5] that a 33° forward slope is most suitable for handle operation. The better performance of the handled trackball at the forward slope of 30° than the forward slope of 15° or 0° can be partly attributed to the fact that the muscle of the wrist and arm relaxes easily at the forward slope of 30° and the stress on the muscle of the thumb, wrist, and arm becomes greater with reduction of the angle. Since the muscle becomes tighter with increase of stress, pain and fatigue are brought about as a result. Accordingly, the forward slope of the handle is decisive for comfort of the user when operating the handled trackball.

Analyses also suggested that the forward-lateral slope combination of 30°-30° were a good design choice. First, this slope configuration reached the shortest click time in the cursor movement task. Second, the usability questionnaires show that the lowest perception of "hard to control this button" appeared at this particular posture. Lastly, both the subjective fatigue and usability evaluations showed that this configuration was among the lowest fatigue and unusable group in several questions. All of this suggests the forward-lateral slope combination of 30°-30° should be used for the design.

This study included the table trackball as a comparison. For the table trackball, the highest fatigue rating was given to the right palm, probably because more forearm pronation combined with thumb and finger movement leads to more exertion. Also, the palm of the right hand is completely placed on the device and thus cannot grasp it smoothly in comparison with ordinary mouse devices. Consequently, the index finger and the palm cannot relax easily and can suffer from undue stress and fatigue. Since most participants never used the handled trackball, the study initially suspected that the performance or subjective impression might be affected by such instrumentation. The study therefore used the table trackball as a control, so that the performance of the new device can be judged and compared with the control. Based on the results of the study, both the performance and subjective rating of the handled trackball are comparable with the usual table trackball. But the handled design offers several design characteristics to allow the user to securely grasp the handle while simultaneously operating the trackball and buttons. When the operator is subject to severe motion, such design provides better stability for the hand and arm posture. Future experimentation is planned to test the design under motion (vibration) environment.

5 Conclusion

The study designed a handled trackball aiming at providing better grasp and stability for the operator to perform computer tasks while in motion. Several handle slope configurations were tested and evaluated for their cursor movement performance and subjective rating of fatigue and usability. The study found that the forward slope was the dominating factor in the handle posture design. The performance of the handled trackball at the forward slope of 30° was better than that of 15° or 0°. The lateral slope had no significant effects on task performance, subjective fatigue or usability. The result also identifies that the handled trackball had the best performance when it was operated at the forward-lateral slope combination of 30°-30°. There were no apparent shortcomings when evaluated by subjective fatigue and usability characteristics. The handled design provides the user with firm grasp while simultaneously working on the trackball. This allows the operator to overcome unstable hand and arm postures disturbed by motion when the device is placed on a ship or motored carrier. Accordingly, this study suggests substituting the handled trackball with the forward-lateral slope combination of 30°-30° for the table trackball to improve the performance of the operation under motion environment.

References

1. Accot, J., Zhai, S.: Performance evaluation of input devices in trajectory-based tasks: An application of the steering law. In: Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, pp. 466–472 (1999)
2. Brumfield, R.H., Champoux, J.A.: A biomechanical study of normal functional wrist motion. Clinical Orthopaedics and Related Research 187, 23–25 (1984)
3. Graham, R.: Motion-induced interruption as ship operability criteria, Naval Engineers Journal, pp. 65–72 (March 1990)

4. Harvey, R., Peper, E.: Surface electromyography and mouse use position. Ergonomics 40(8), 781–789 (1997)
5. Hsu, S.H., Cheng, Y.H.: The establishment of ergonomic design guidelines on non-powered hand tools. Journal of Occupational Safety and Health 6(1), 101–111 (1998)
6. Hsia, P.T., Drury, C.G.: A simple method of evaluating handle design. Applied Ergonomics 17(3), 209–213 (1986)
7. Kong, Y.K., Lowe, B.D.: Optimal cylindrical handle diameter for grip force tasks. International Journal of Industrial Ergonomics 35, 495–507 (2005)
8. Putz-Anderson, V.: Cumulative trauma disorders: a manual for musculoskeletal diseases of the upper limbs. National Institude for Occupational Safety and Health, Cincinnati, Ohio, USA (1988)
9. Thomas, G.W., Henry, H.E.: Effects of Angle of Approach on cursor Movement with a Mouse: consideration of Fitts' Law. Computer in Human Behavior 12, 481–495 (1996)
10. Woodson, W.E., Tillman, B., Tillman, P.: Human factors Design Handbook. McGraw-Hill, New York (1992)

Impact of Culture on International User Research -A Case Study: Integration Pre-study in Paper Mills

Anna Oikarinen[1] and Marko Nieminen[2]

[1] Helsinki University of Technology, TKK
Department of Computer Science and Engineering
P.O.Box 5400 (Konemiehentie 2) FI-02015 TKK Finland
[2] Software Business and Engineering Institute
P.O. Box 9210
FIN-02015 HUT, Finland
{Anna.Oikarinen,Marko.Nieminen,TKK}@tkk.fi

Abstract. Global paper industry needs systems that can be used in all locations. International user studies can be helpful when integrating systems. Not only due to the lack of common language but also the differences in culture and the usage of systems, information from different countries needs to be collected and analyzed so that the integration development is not be biased and unilateral. During the study some food for thought was gathered on what to consider when planning an international user study.

Keywords: International user study, integration, Hofstede, cultural theory.

1 Introduction

When planning an integration project, ABB wanted to determine what would be the ideal solution from customer's point of view in integrating systems and decided to organize a user study with its customers. The users were mainly machine operators and management personnel in paper mills. The user study started in Finland but soon the stakeholders decided that it should be international. As a result, an international user study was planned to take place in Finland, Spain, Indonesia and China. The methods used were interview and observation in the paper mills.

1.1 Background

Most of the earlier scientific international usability research is about how to keep culture in view when planning user interfaces. General guidelines are given in texts such as Culture and Design in International User Interfaces [1] and Global Interface Design [2]. There are also some articles about conducting international user testing, but not so much on doing user research internationally. The subject is dealt with in seminars like Evaluating Globally: How to Conduct International or Intercultural Usability Research [3], Techniques for Researching and Designing Global Products in an Unstable World: A Case Study [4], and Managing International User Research [5]. Dray has done several seminars, conferences and some articles about international

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 576–585, 2007.
© Springer-Verlag Berlin Heidelberg 2007

user research. However, there are no notes or articles available on those seminars, so the knowledge is not available to all.

So the reason for this article is the lack of available literature on the subject of what kinds of effects culture has when studying users. In this study, a less specifically user-oriented culture theory of Hofstede was chosen because it is widely known; it covers several cultures and includes the effect of culture on work behaviour [6].

1.2 Aim of the Research and Research Question

The cultural aspect of conducting usability studies is significant. The problems caused by cultural differences and the lack of common language were collected as they rose, and the aim was to describe them as well as conducting international user studies in industrial environment, i.e. in paper mills. The aim of the study was not to compare cultures but rather discuss about the differences of them based on Hofstede's theory. The main idea was to present the results as some examples in order to understand the differences in cultures under study.

2 Cultural Theory

In order to discuss the differences of culture it is vital first to define what culture is. The behavior, knowledge and beliefs of humans are intertwined in a package of culture that is learned and transmitted to the next generations. Culture affects everything in one's life: "An individual's attitudes, values, ideals, and beliefs are greatly influenced by the culture (or cultures) in which he or she lives." [7].

In this kind of research, where one person is interviewing and observing and the other person is subject to it, culture is mostly evident in the interaction between the people. The biggest cultural gap is between the sending of a message from one culture and interpreting it in another. Here are the five dimensions of culture by Hofstede shortly presented.

Power Distance: People behave differently towards the inequality they see in the society. Some are accepting it as a way things should be, whereas some do not tolerate it at all. Hofstede measures the "dependence relationships in a country" with a power distance index (PDI). The smaller the index, the more likely the people are non-dependent on their superiors and prefer consultative leadership, and the equality in the workplace is valued. Young and hungry bosses are admired more than old hacks. The larger the index, the more dependent the workers are on their bosses, and more likely the organizations have very hierarchical organizations. Respect is gained with age and experience. [6]

Individualism vs. Collectivism: Individualism can be described as an attitude that allows putting one's own interests over the interest of one's group. In work, individualist seeks for personal time, freedom and challenge. Better salary or benefits are completely acceptable reasons for changing employers, which would never apply in collectivist cultures. Everything needs to be expressed explicitly verbally, and silence is considered abnormal. [6]

In collectivist cultures individuals put group's interests over their own. They seek for training, physical conditions and use of skills from their work. Face is a mental model of a fair relationship with the social environment and it is as important as the person's actual face. The loosing of face does not only apply to oneself but also to others, so that is the reason why direct confrontation is regarded as extremely rude. So for example no is replaced with expressions such as "you may be right" or "we will think about it". Loyalty towards family and employer are important. The resources are shared and for example salary is given to the family. Communication relies on the context and non-verbal communication rather than direct expressions. [6]

Masculinity vs. Femininity: In masculine societies the men and women differ in their values: men are tougher and more competitive and women more concerned with life quality. The more masculine the society is, the bigger is the difference between men and women in this respect. Also the values of women harden, but not as much. Both genders embody both of these qualities, only the amount differs. Masculinity refers to what is being considered as important in work: things such as good wages, recognition, advancement opportunities and challenges; it highlights competition and imposing oneself. Conflicts are settled by flaming arguing, and the role of work is important; they "live in order to work". Reward comes to those who deserve it. [6]

Femininity refers to a society where these social gender roles are partly overlapping. It is associated with having good relationship with one's boss, living in good environment and having secured post. Values associated with it cover modesty and unambitiousness. One should not appear too eager: the common solidarity is the accepted goal. Arguments are settled with compromises and mediation. Work is just something one does in order to afford living. People at work are rewarded based on equality. [6]

Uncertainty Avoidance: The tolerance for ambiguity and unpredictable is one of Hofstede's dimensions. It is "the extent to which the members of a culture feel threatened by uncertain or unknown situations" [7]. Hofstede uses term anxiety as describing the feeling that comes from uncertainty. "Anxious people" are more expressive in their gestures, they use their hands in explaining things and raise their voices, and they express their feeling more openly. However, expressing emotions and aggressions is not allowed in some cultures due to the social norms so stress cannot always be demolished in action. [6]

Avoiding uncertainty leads to a search for predictable and structured life. Strong uncertainty avoidance cultures have strict objectives, timetables and detailed tasks. The work place has formal and informal rules that one has to follow in order to fit in and be accepted. Rights and duties lead to a controlled situation; people have emotional need for structure. In weak uncertainty avoidance cultures people pride in solving problems without distinct rules. Even though they do not have as much norms as the strong uncertainty avoidance cultures, they tend to respect the ones they have more. [6]

Confucian Dynamism vs. Long-term Orientation: Hofstede's fifth dimension can be found from the Chinese philosopher Confucius teachings about "practical ethics". These teachings are not related to religion but rather practical advices on how to live one's life. Long-term orientation in life includes the qualities such as persistence,

economy, feeling of shame, defining relationships based on status and observing that order. People want to adjust traditions into the current conditions and they save money. They are willing to spend time in order to get something and they want to respect the Virtue's demands. It is oriented towards the future, unlike the short-term orientation that is focused on the past and the present. It includes the qualities like personal solidity and staidness, respecting traditions, protecting one's face and responding to gifts, compliments and favours. People compete socially whatever the costs; they have usually a little money to spend and they want quick results. [6]

3 User Study in the Paper Mills

After the theoretical frame of Hofstede's dimensions is presented, it is important to define, how the study was conducted. Six paper mills were studied in four countries: three paper mills in Finland, one in Spain and Indonesia and two in China. The paper mills were owned by different organizations, and the size of the mill (personnel, paper machine number etc.) as well as the paper type varied. The personnel varied from 800 to more that 13.000 and the amount of paper machines from 2 to 12. The titles and hierarchy between different countries varied as well as the scale of responsibility areas, but the tasks were rather similar independent of the paper mill. The positions under study included shift manager, production manager, machine operator, proportioner, development engineer and so on. Total number of interviews was 36: 15 in Finland, 7 in Spain, 3 in Indonesia, and 11 in China. The final products of paper mills covered LWC, art printing papers, corrugated paper, coated and non-coated fine paper, office paper, and printing paper.

Next, the methods and the analysis process will be shortly described.

3.1 Methods

Methods used were interviews and observations done by two women. Originally team used the Contextual Design-methodology [8]. But since it became evident that CD as such was not suitable for this kind of study due to the language and time constraints (more in chapter 5), methodology was changed. The main difference between using CD versus using interview and observation is first of all the analysis phase and second, that in CD interview and observation are done overlapping, but in this case they were done sequentially, and in some cases, even with different people.

The interview method used is between semi-structured and theme interview due to the rather obscure subject. The interviews took place mostly in the user's work environment but they were asked predefined questions and follow-up questions to relevant issues that were noticed. [9]. Main focus was on the tasks and needs of users regarding different systems and information related to users work. The questions were related to the tasks, work, information needed and systems used. In addition, in Finland the operators were both interviewed and observed, whereas in other locations the language level was so poor, that interviews were impossible, thus leaving observation as the only reasonable possibility to gather information. The observation lasted usually from half an hour to an hour. During this the goal was to observe the use of different systems, the sharing of information, and the role each system had in

the control room. All visits except one done lasted two days; one visit lasted only one day due to traveling schedule. Overall, two days was better for gathering the information than just one day, and even more could have been useful. Most of the information gathered abroad was done through interviews.

3.2 Analysis

Reporting and analyzing cultural material is not very structured process. As the material is clearly qualitative, numerical analysis is not possible. The results were reported as interview and observation notes and summaries written by the researcher. Some user descriptions were written based on the responsibility areas and tasks as well as the different systems they use. Rather than specific and structured instructions the results will be presented as case examples of situations encountered during the study. The behaviour presented here is not universal but the situations can occur in different circumstances, and correct behaviour can be adopted and applied according to the cultural dimensions.

4 Experiences on Interviews and Observations from a Paper Mill Study

In this chapter the differences encountered will be discussed by country. To link the dimensions of Hofstede to these examples, a short explanation of the countries position compared to Finland will be given. This of course does not apply to Finland in which case the overall position compared to the world average will be described. The values for each country are presented in Table 1.

Table 1. Hofstede's dimension values

Country	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	Confucian Dynamism
Finland	33	63	26	59	-
Spain	57	51	42	86	-
Indonesia	78	14	46	48	-
China	80	20	66	30	118

4.1 Finland

Finland scores pretty low in the power distance index. The world average (WA) is about 60. So Finland's low PDI means that the power is distributed equally and if it would not be, it would not be accepted. Individualism is above WA which is 40. Finnish people take care of their own family and seek for challenge. Things need to be expressed rather explicitly and honesty is valued. Masculinity is also pretty low compared to the WA. People value quality of life and gender roles are not so distinct. Uncertainty avoidance WA is about 64. Finnish are not very accustomed to express their feelings openly, but rather have a tendency to keep their emotions to themselves. There is no information about the long-term orientation. [6], [10].

In Finland the language and culture were familiar. However, there were some macho attitudes towards women conducting the study. In the control room tasks were clearly separated. The role of machine operator was the most responsible one, since he was "in control of everything related to the machine itself". The bosses and subordinates had equal relationships. They worked on the last name basis, which in Finland is a positive thing. The subordinates were not afraid to say their opinion and they used this option often. They were listened to and their opinions respected. Experience based knowledge was more valued than education, and so the responsibility grew as experience was gained. There were no clear rules for each situation but the operators rather things the way they were used to. Own perception of the situation was valued more than the information systems gave and so the decisions were often based on that. There was no clear respect for the authorities, but instead a strong commitment to co-workers. They knew each others hobbies and families and spent time outside work.

Operators have to be able to stand uncertainty, and they usually take the situations as they came. The world in the mill is masculine based on the physical environment of heavy machines. Strength is valued and respect has to be earned; it is not given based on fancy cars or titles. Young and well educated manager is not as respected as an older manager who has started from the bottom.

4.2 Spain

Spain has bigger PDI than Finland. It scores lower in individuality, but higher in masculinity with almost doubling of Finland's value. When the distribution of power is unequal, it is not accepted as willingly. Spain is a less individual country, with respect to taking care of one's family. Spanish are also more masculine than Finnish meaning that they value masculine qualities such as competitiveness, assertiveness and recognition. It has stronger uncertainty avoidance than Finland. People are very tolerant to ambiguity; they are more expressive, use their hands and raise their voices, unlike the Finnish, who are withdrawn in those aspects. They do not have many strict rules but when they do, they tend to follow them. [6], [10].

Spain is traditionally seen as a macho-culture and this affects the perceptions made in there if not realized. As a supporter to this image of women staying at home, majority of the workers in the mill were men. But the women that worked there were respected based on their skills. For example one woman was respected because she was an excellent lab technician and knew more about the chemical side of the paper making than the men. Education is one key to respect. Even young women are respected if they are well educated, ad they can get a good position in the organization. The respect and the position are earned; they are not given. Education is just one way to "earn" them. The upper management is an exception, their possessions gather respect. The employees in the mill were more educated than in Finland, but the language skills were not so good. Most of the operators did not speak English: the ones, who could, acted as interpreters.

Spanish people used a lot of gestures while speaking and they generally raised more their voice. The co-workers talked to each other constantly and gave each other advices by showing how to do things. They also were very polite towards women: they wanted to open the doors and carry the bags. However, this was not the only way

the politeness was shown: they also wanted to give a good impression of their skills in using the system and their usage rates. In a way they wanted to make the supplier feel good about the meeting. They also had a guide for us all the time and they tried to fulfil all of our wishes and needs. They wanted to know about our country and our origins as well.

The bureaucracy related to taking pictures was enormous: after one hour of waiting, the situation was settled by requiring passport number for documentation. The permission was given, but the request of taking photographs caused somewhat a situation that was not predicted. But the Spanish handled the situation fluently, even though they did not have existing rules to follow.

4.3 Indonesia

Indonesia's PDI is twice as large as Finland's index. It means that the power is unequally distributed but that this inequity is accepted by the society. I t is highly collectivist country, so there is a strong commitment to family and the in-group. Direct confrontation is undesirable and very little needs to be expressed verbally. In Indonesia responsibility of the group members is taken by everyone. It is a more masculine country than Finland. Uncertainty avoidance is 48, which makes Indonesian create strict rules, regulations and laws in order to minimize the ambiguity. They do not accept change easily and they try to control everything. [6], [10].

Indonesia presented a culture shock when first arriving. The paper mill was dirty and it was nothing like previous mills visited. The group had enormous language problems even though they had a local ABB representative with us, who spoke the local language, as well as a Finnish male companion, who had been to the mill before and knew the people. However, this was not enough. The group did not share common language and the local men were shy of them, the questions were not understood, and if they were understood, the answers were not understood. However, the group did manage to gather some information by trying to be open and by reading the signals sent by the locals. Finnish way of dealing was not suitable for interview in Indonesia. They wanted to answer as a group and presenting the question to one person was not a vice decision. The oldest member of the group had the most power and wisdom, so the others respected him and ultimately the answer was his. There were some negotiations but the authority of the oldest made sure, that his opinion and expertise was taken into account. Others were not eager to express their ideas, especially not when directly confronted. The people wanted also everyone to keep face as well: they did not let situations where someone might loose his face, occur.

4.4 China

The biggest difference between the dimension values is with Finland and China. The PDI of China is similar to Indonesia in the equality of the distribution of power and its acceptance. China is also similar to Indonesia in the collectivism, meaning that Chinese also have strong and committed relationships, and they take care of each other: very little needs to be said due to the so-called high-context culture. China is a rather masculine culture meaning that it is competitive and masculine values are

endorsed. The uncertainty avoidance index tells that the Chinese tolerate uncertainty very poorly and that they have created rules and laws to reduce its effect in society and life. Long-term orientation is the highest measured, and it applies to all Asian countries. They want to go through the rock with strength, will or time. [6], [10].

In China one of the mills was owned by a Finnish paper company and the other by a Chinese one. The differences were significant. One had Finnish upper management and the language used was English. The management used western style and they had to use middle men to pass the orders. The goal was to bring the good work practices and finally give the control to the locals, so the decision making done only by Finnish was seen as a problem. The interviewees' language skills were good and they were willing to help, and they answered the questions willingly. The working culture was similar to the ones seen in Finland. The welcoming was warm and the people pointed to help took their job seriously. The operators were quieter and the lack of common language resulted in a situation, where mostly the ones with good enough language skills (namely people in positions between management and operators) were interviewed. They respected their superiors and they seemed loyal to the employer. Education level was higher than in Spain or Finland: many of even the operators had a university degree. Actually, everything went as the group would have been in Finland: even permission for taking photographs was granted easily.

In the other paper mill the situation was completely different. The language skills of the personnel were rather poor, since the language used was Chinese. The attitude of the mill manager was tough towards the research group: he had no time to look after them so he just abandoned them with one of local ABB workers to wait for someone to arrive. In the interviews, there were several people present. The guide of course acted as an interpreter, but did not succeed so well due to insufficient language skills. Trying to ask the same question in several different ways, the guide and the interviewees discussed for five minutes or so, and finally the answer was given by the guide. His expertise of the subject made him discuss about the answer with the interviewee, so as a result, the team got as an answer the final result of the discussion between two people. The hierarchy proved out to be a problem. If the operators would have answered, the guide would have lost his face. The guide, who was pointed in the job by the mill manager apparently on a very short notice, seemed stressed about the duty. He tried to fill group's wishes but the bureaucracy of the mill came in the way. The group could not take photographs, it was strictly prohibited. The group had to follow certain routes in the mill, and visit the control rooms in certain order. They could not go to the other machines because it was not agreed upon beforehand. The loyalty of the workers was extremely high: many of them had been working in the other mills of the company and transferred to this new mill.

5 Discussion of the Methods

In Finnish process industry, Koskinen has done a lot of research especially in the paper mills with CD-methodology [11], [12], [13]. Even though the results have been good, in this case the methodology could not be fully benefited from. The CD-methodology was chosen because it was familiar to the organization, but the main problem in this case with CD was that it seems to be designed for office environment.

Granted that the modern control room is similar to office, the tasks done in it are not. The main difference about the work itself is described well by one of the interviewees: “In an ideal situation, you don’t have to do anything”. Operator’s work is mostly related to paper runnability and they do not have constant duties. Many problems can be found out before they become actual problems based on the experience as well as the information provided by different tools. Due to the fast pace of the study (maximum of two days per location), the material collected this way did not answer the questions and something more specific needed to be done to benefit the time available to its full potential.

The fact that this study was conducted in several countries, is one of the most critical things that the CD does not take into consideration: different cultures introduce new problems as well as the language barriers. The observation can be of course done but the interviewing demands an interpreter, which again presents different kinds of problems. The interpreter needs to be familiar with the subject in order to know the correct terms and excel in both of the languages used. However, he cannot be too closely involved with the tasks, because the knowledge and experience makes him “filter” the information. An example of this is the case in China, where the use of interpreter did not succeed. And unfortunately, the observation without interpreter leaves the information somewhat superficial without the knowledge that the interview might add up. This means that the information gathered in different countries might not be as comparable as hoped beforehand and it also might be biased.

6 Conclusions

Even though Hofstede’s theory is mostly used in studying cultural differences, it can also be used for preparation of international user research. Also the fact that Contextual Inquiry needs to be adapted in this kind of research, can be useful when planning international studies. One cannot say anything fundamental of the Hofstede’s dimensions based on this study but they helped in preparing for the journeys and when adjusting to the situations. The theory was used rather as a context provider than an ultimate truth of what people are like; for giving guidance on what kind of presumptions should be tackled with and what in ones own behaviour should be modified. Very deep analysis of the differences of the culture could not be done but some insight on relationship between behaviour, politeness and the adaptability. So as a result there are no exact instructions but rather some food for thought.

Acknowledgments. I’d like to thank the following people for their support in writing this article: Erika Salmela for being an excellent research partner; other WIS team members for everything; Marko Nieminen for encouraging me to write this article and giving valuable comments; and finally both Emilia Oikarinen and Juhani Mykkänen: without them this would never have been completed.

References

1. del Galdo, E.: Culture and Design in International User Interfaces. Wiley Computer Publishing, New York, NY (1996)
2. Fernandes, T.: Global Interfaces Design. AP Professional, Boston, MA (1995)

3. Roshak, L. (organizer), Spool, J. (moderator).: Evaluating Globally: How to Conduct International or Intercultural Usability Research. Panel discussion. CHI 2003 (April 5-10, 2003) FT Lauderdale, Florida USA (2003)
4. Foucault, B.E., Russel, R.S., Bell, G.: Techniques for Researching and Designing Global Products in an Unstable World: A Case Study. CHI 2004 (April 24-29, 2004) Vienna, Austria (2004)
5. Dray, S., Mack, A., Larvie, P., Lovejoy, T., Prabhu, G., Sturm, C. (2006) Managing International User Research. Panel discussion. CHI 2006 (April 22-27, 2006) Montreal, Quebec, Canada. Available online at http://delivery.acm.org/10.1145/1130000/1125454/p5-mack.pdf?key1=1125454&key2=5424731711&coll=GUIDE&dl=GUIDE&CFID=14265755&CFTOKEN=79577815
6. Hofstede, G.: Cultures and organizations: software of the mind. McGraw-Hill, London (1991)
7. Merriam Webster OnLine Dictionary. Definition of culture. Available online at http://209.161.33.50/dictionary/culture
8. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann, San Francisco, CA (1998)
9. Huotari, P., Laitakari-Svärd, I., Laakko, J., Koskinen, I.: Käyttäjäkeskeinen tuotesuunnittelu. Käyttäjätiedon keruu, mallittaminen ja arviointi. Taideteollisen korkeakoulun julkaisu B 74. Gummerus Kirjapaino Oy, Saarijärvi (2003)
10. Hofstede, G.: Geert Hofstede Cultural Dimensions (2003) Available online at http://www.geert-hofstede.com/hofstede_dimensions.php
11. Kontio, K., Nieminen, M., Koskinen, T.: Rauhamaa Lightweight Contextual Design: A case study in process control product development. 10th International Conference on Human - Computer Interaction. Crete, Greece (June 22-27, 2003)
12. Koskinen, T., Nieminen, M., Paunonen, H., Oksanen, J.: Process Snapshots Supporting Operators' Expertise Management. 10th International Conference on Human - Computer Interaction. Crete, Greece (June 22-27, 2003)
13. Koskinen, T.: Understanding the Control Room Context: Defining Requirements For Computer Supported Knowledge Sharing. In: Proceedings of the 34th NES Congress Humans in a complex environment - Innovate, integrate, implement (2002)

Computer Mediated Banking: A Cross-Cultural Analysis of SMEs

Alison Ruth and Jenine Beekhuyzen

Griffith University, Brisbane, Australia
{a.ruth,j.beekhuyzen}@griffth.edu.au

Abstract. This paper presents a view of banking as undertaken by SMEs (Small and Medium Enterprise) in Australia. It presents a user perspective to give insight into how people talk about banking, how they are using traditional bank services, and what it means to them to bank with new technologies. This paper builds on previous analysis and interpretation of the perceptions of these issues in the banking project. In this paper we apply Burke's [5] dramatistic analysis. The paper analyses 15 SMEs to elaborate the mediation of money between banks and individual SMEs. We found that when talking about banking, individuals refer to location (scene) and processes using cheques, cash and the online interface (acts and agency) Thus an elaboration of the elements indicates that the scene-act-agency interaction is perhaps a significant nexus through which individuals negotiate this activity.

Keywords: SME, banking, Internet.

1 Introduction

The activity of banking is complex. With new technologies come new ways of transacting that have a huge impact on small business. The complex nature of banking requires a variety of approaches to understand this complexity. Previous analyses have been presented on our data; from an actor network theory perspective [2] to a grounded theory approach [16]. These perspectives have provided great value in understanding banking activities, but have also added some complexity through allocating the computer the status of agent. The alternative but complementary view presented in this paper places the computer as a part of the process of enacting banking. Hence here we present a new perspective of banking using Burke's [5] Pentad as the method of analysis, and this view of banking as a communicative act helps to identify issues relevant to cross-cultural design.

Small and medium enterprises (SMEs) in the global environment provide the context for this paper. The literature suggests that SMEs need to be considered differently to larger organisations in terms of technology development, adoption and usage. Indeed most major international banks are developing internet banking specifically for SMEs although cash may still be the dominant form of payment for SMEs worldwide in the foreseeable future [12]. The reason for this is that technology can have a huge sociological and economical impact, as well an important cultural impact [10]. This paper concludes that SMEs are a distinctive cultural group. We

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 586–595, 2007.
© Springer-Verlag Berlin Heidelberg 2007

build on our previous findings of user-centred security where we have found that in order to enhance banking security we need to not only focus on the technical aspects, but also the social aspects of design. It is important to enhance trust; and this can be done by increasing ease of use, people's control and personalisation of information, and showing care for the customer [15].

Australian SME's (and many others worldwide) are known to be slow to adopt new technology and to use technology innovatively. Somewhat contrary to this, our study shows that Internet banking is being embraced by SMEs. It is providing flexibility of work practices and the ability to control various aspects of business and personal banking activities (where, when, how); previously this was less possible.

Burke's Pentad provides specific attention to the location in which an activity takes place (where), as well as the individuals (who), the means of mediation of the activities (how), the activities and interactions that occur (what) and the purpose for which the individuals are participating in the activities (why). In this way, it goes beyond many current theories that emphasise situational factors. That is, by including the kinds of interactions which not only occur in actively constituting a banking environment, this framework represents a more encompassing explanation of banking environments than those privileging the physicality or the activities of the particular setting. This holistic approach provides insights that may not be available using other frameworks.

2 Theoretical Basis

This paper synthesises data from interviews with individuals and integrates it with the conceptual framework of Burke [5] into a unified and cogent conceptual basis for describing and evaluating a mediated activity such as banking. Burke's [5] dramatistic analysis using the Pentad is a valuable methodological tool for investigating communicative acts [13, 14]. Ågerfalk and Eriksson [1] previously approached the interaction between an individual and an ATM as a communicative act using Weber [19] and Habermas [8]. While their analysis focused on usability, particularly effectiveness, efficiency and satisfaction, they argue that usability must be approached with consideration of both the instrumental and the communicative goals of a system. Like Ågerfalk and Errickson, we are concerned about how an individual interacts with a bank within a social system even if 'the user is potentially unaware of or uninterested in this larger social context' [1: 10].

2.1 Description of the Pentad

Burke [5] proposes five terms or pentadic elements (act, scene, agent, agency and purpose) that assist in understanding a narrative, in this case banking. Using this framework allows focus to be placed successively and simultaneously on elements within the pentad. This ensures that any discussion of banking (what is done – an act) will proceed from focusing attention on the scene, where people are banking, in tension with one or more other elements, who they are, how they are doing it or why. These elements contribute to our understanding of the interactions banking customers engage in and how we can design our banking systems to make the most of these

interactions in terms of security, privacy, identity and trust. Burke [5] states that any analysis of the motives of participants will give "*some kind of* answer to these five questions: what was done (act), when or where it was done (scene), who did it (agent), how he did it (agency) and why (purpose)". (p xv italics in original).

While each element can be used to describe in detail one aspect of what we refer to as the 'banking environment', Burke elaborates this through the ratios that are manifest between the elements. These dyads, for instance, the scene-agent ratio, provide greater insights into what motivates people to undertake an activity, because it is through the elaboration of the tensions between the act and the scene, for instance, that the meaning of the act is further illuminated. Each element, even when discussed individually, that is, through attention to each term as a single element, still requires other elements to assist in its definition. In this way, the understanding of motivations is illuminated, because the purpose for engagement, for instance, is highlighted by how individuals engage with the scene (agency).

In terms of a banking environment, some acts have no form or possibility outside the environment. For instance, depositing money requires a bank to accept the money for holding in an account; else it becomes a gift or a purchase. Thus, a customer may engage in the act of banking within a bank, at an automatic teller machine or online, each act constituting banking, but each constrained by the kinds of act available through the 'interface', that is, the 'teller', the ATM or the computer. A customer in a bank, interacting with a 'teller' has access to more information and more forms of interaction than a customer at an ATM. Similarly, a customer using internet banking has access to more information than that available at an ATM, but may be constrained by the location of the computer in a 'public space' (e.g. an internet café). A customer interacting with Internet banking within a more private location will be less constrained than one interacting in an Internet café (a more 'public' private space yet functionally equivalent in terms of the internet banking interface).

3 Research Methods

Presented here are results of a qualitative study on banking, personal communications and financial decision-making. This research is part of a wider project focusing on Security, Trust, Identity and Privacy in the Smart Internet Technology Cooperative Research Centre. We approach these issues from the users' perspective making them central to the design of financial services and bank strategy.

We conducted a qualitative study between April 2005 and March 2005, with fifteen participants from SMEs across two capital cities in Australia: Melbourne and Brisbane (see Table 1 for details of the characteristics of participants, (all names are pseudonyms)). The people were accessed through personal and professional networks. In choosing our sample for the qualitative interviews, we were careful to include bank customers who do engage in online banking as well as those who don't. We adopt a user-centered design perspective which places users' activities within their social and cultural context [18, 3, 17]. It is a 'grounded' study in that there was a fit between data and emerging theory, rather than a testing of hypotheses [7].

Table 1. Participants' characteristics

Name	*AGE*	*Household income*	*Uses internet*	*Ethnicity*	*Type of SME*	*location*
Brenda	35-44	75-99	Yes	Australian	Farmer	Rural
Claire	45-54	25-49	No	Australian	Farmer	Rural
Fran	35-44	75-99	Yes	Australian	Graphic Design	Other Urban
Fred	45-54	$100k+	No	UK	IT services	Urban
Gay	25-34	100+	No	Australian	Medical	Urban
Greg	18-24	Under 25	Yes	Greece	IT Services	Urban
Hester	55-64	$100k+	Yes	NW Europe	Biotechnology	Urban
David	45-54	100+	Yes	Australian	Consultant	Urban
Laura	25-34	N/A	Yes	Australia	Medical	Urban
Mark	35-44	$100k+	Yes	Australian	Consultant	Urban
Nancy	35-44	$75 - 99	Yes	Australia	IT Services	Urban
Peter	25-34	$100k+	Yes	Australia	IT Services	Urban
Rita	35-44	$100k+	Yes	SEAsian	Unknown	Urban
Samuel	35-44	100+	Yes	Australian	Hospitality	Urban
Shane	25-34	N/A	Yes	Australian	IT Services	Urban

Each face-to-face interview lasted between one and two hours, being tape recorded with permission from the participant. The interviews were then transcribed. The project team used QSR's N6, a computer program to assist with qualitative analysis. The use of this software helped to keep all project data in one location and accessible, as well capturing themes, thoughts and general ideas arising from the data during analysis. Each team member was involved in the coding of interviews. We approached this by first broadly coding the data, then organizing the data into matrices to check emerging themes in a transparent manner. We also used the N6 software to identify negative cases so that the study was rigorous.

SMEs from a range of industries were interviewed for this study. All participant SMEs contributed well to the diversity of the overall project by giving useful and relevant information about personal banking activities as well as those that are business-oriented. The separation between personal and business banking was often difficult for those running an SME.

Our analysis reveals how a particular group of banking customers (SMEs in Australia) constitutes a distinctive cultural group. Comparisons among banking environments and individuals participating therein are facilitated by a Pentadic Analysis, because each example may be viewed through the same lens. These lenses can be applied to different situations and different cultural groups providing clearer evidence of which form best supports banking for a particular group of customers.

This research constitutes the first of a series of analyses of banking environments which all differ in their implementation and delivery. However, the speed with which technology evolves means an increasing need to establish those aspects that do facilitate secure banking in these different situations, and incorporate them into the design of banking systems designed for particular cultural and business groups.

4 Usability of Banking Interfaces

Ågerfalk and Eriksson [1] found that within a communicative orientation, mutual understanding and trust were key interpretations of usability criteria. These concepts relate to security, privacy, identity and trust as the mutual understanding requires that the user and the banking system require security to ensure the identity of the individual is correct and that privacy is maintained to prevent the erroneous allocation of an identity to another individual. Thus, using Ågerfalk and Eriksson's framework with Burke's framework allows the use of Burke's Pentad to elaborate these key concepts of mutual understanding and trust.

4.1 The Elements

Each of the five elements interacts with other elements in many ways resulting in many interrelationships between elements. These interconnections depict the dyadic relationships that arise from the analysis. Each of these elements requires viewing from multiple perspectives to understand their contribution to a banking environment. The interplay of the dyads illuminates how each element contributes to the definition of banking and associated activities.

Agent The agent in this context is the account holder or a designated person who can act on their behalf. In this research, it is the SME and this is generally made up of an individual person or a partnership (often husband and wife).

Purpose The purpose of banking is embedded in daily life and business keeping tabs on finances, paying bills. Other purposes come into play particularly with business for managing the business finances and the business in toto. One of the main purposes of banking relates to security of finances and money.

Act The act is a composite process of depositing, withdrawing, investigating balances, paying bills, monitoring credit. Any interaction with a banking institution may be classed as an act, from gathering information to borrowing money and setting up retirement schemes.

Agency Agency has a number of levels. From one perspective, the Bank is the agency through which agents act with their money. The Bank essentially enables agents to act. At a lower level, the branching system of banks in Australia affects the possible acts, and whether an individual agent can access the same information via different branches. Note that some Banks create Agencies (branches) for customers. Individual agency is often limited through these Bank Agencies. In effect, such Bank Agencies deny certain levels of individual agency, that is, how individual agents interact with the bank is curtailed by the Bank Agency. Again at a more confined level, agency is enacted through the forms and processes which the bank uses for exchange of information (deposit forms, withdrawal forms, loan applications etc) as well as the passbook and the magnetic strip card (key cards, credit cards, debit cards).

Scene The scene is another area which has multiple definitions and one which is becoming more complex with the changes in options for individual agents. Initially, the scene is the Bank Office/Premises during office hours (the scene also has a temporal aspect – when particular acts may be undertaken). Prior to the advent of information technologies, individual agents could often only act within their defined Bank Branch. Systems enabled certain actions to be undertaken in the non-primary

branch, but often access to information was limited. Information technologies, including electronic data transfers and the Internet have enabled more fluid interactions with Banks, such that individuals could gain access to all their information at all Branches (notwithstanding that some Bank Agencies still had limited access to information). In this way, the scene was expanded - individuals had multiple options for gaining access to their information.

However, given the mobility of much technology, particular laptop computers, personal digital assistants (PDAs) and mobile phones, the broader definition of the scene needs to include the locality of the device through which the individual agent accesses their information. For instance, there are differing requirements accessing banking information in an office environment than there are accessing the same information in an Internet cafe. The types of interactions can be limited particularly if the screen can be easily viewed by a passerby. Similarly, the operating system of the host computer is often not as secure as an individual's own computer, depending on the nature of the access. These factors increase risks to maintaining security and privacy of information and potentially open up new risks to an agent's identity.

4.2 Communicative Acts

Having defined the pentadic elements in brief within the banking environment, the remainder of this paper will focus on the nexus of banking; that is the interaction of the scene, act, and agency and the dyads arising within this interaction. It should be noted from the above descriptions of the elements, that the agent is defined in terms of SMEs, and the purpose is banking related. Thus two of the five elements remain constant for this discussion. Scene, act and agency vary considerably between forms of banking, the particular activities undertaken and where the activities are occurring. Thus the following elaboration is focused on three elements – the nexus of banking.

5 Patterns of Usage Within SMEs

To contextualize the interactions of the 15 individuals within the SMEs and their relationship to Internet Banking (IB), table 2 demonstrates the comfort with which individuals interact with online finance. While the majority appear to be comfortable with both Internet banking and e-commerce, some prefer not to use IB while one person does not use either. This demonstrates that the scene may be a particular point of conflict for some users.

As highlighted by table 2 (and explained in more detail previously in Section 3), Gay appears to be an anomaly within the SME group under study. She does not use IB or ecommerce. She deposits the cheques for her business in the bank. This demonstrates that her acts (depositing) are undertaken with a particular agency (cheques) in a particular scene. Other forms of agency used by Gay include cash, which she spends or puts away in her cash drawer. Her interaction with the bank (scene) with cash (agency) is confined to depositing cash (act-agency) when she knows that a bill is due (purpose). Gay has two credit cards, but uses them only when she has to pay for something on the phone. Otherwise it is cheques, EFTPOS or Bpay

over the phone, cash for small items. In previous work situations, she received a cheque for her pay, but otherwise it was all direct credit. Now in her business, she receives mostly card payments, some cheques and some cash. Gay doesn't use the ATM (agency); she gets cash from work (scene-act-agency).

Table 2. Participants' preferences for IB and e-commerce

		Internet Banking	
		Yes	No
E-commerce	Yes	Peter, Fran, Hester, Mark, Nancy, David, Samuel, Shane, Rita, Laura, Greg	Fred, Claire
	No		Gay

Fred still uses cheques (agency) stating that he is 'still old fashioned enough' to do so; although he goes on to say he is 'using them less and less'. Fred likes to say hello (act) to the people in the bank (scene) and goes once a week to deposit cash (act-agency). Fred feels that in an age of internet transactions, Bpay and direct debits and credits, the notion of the friendliness of the bank and the personal transactions (acts) that he has are important to him and he wants to know that 'there is someone I can go to who knows about me' (agent-agency). Fred's way of interacting with the bank displays an awareness of the locality (scene) of the acts he undertakes and provides a sense of individual security for him so his role in his banking (agency) empowers him and maintains his sense of identity. Fred states that IB is 'not convenient'.

Claire, likewise, did not like IB, feeling it was unsafe. Claire only uses phone (agency) banking for balances. She may eventually use IB (scene) more but is concerned about security. She states 'it's the not knowing' about security that has prevented her from accessing more services.

The remaining individuals show varying levels of individual interactions with their banks. Peter and Hester visit the bank less and less (act-scene), Hester because it is 'convenient' and Peter because his business is now chiefly with governmental departments. However, Peter states

> I still receive cheques (agency) from other people particularly from smaller clients (act-co-agents). Small business' (co-agents) that have ..only twenty or thirty people or down to sole trading. And very, very rarely people pay in cash, extraordinarily rarely.

His income from the government work is by direct payment (act-agency). Peter BPays 'just about everything' (act-agency-scene).

Hester's business on the other hand, deals with large sums of money in an international context and uses a specifically designed system (agency-scene). Hester spends much less time at the bank (scene) now he has Internet banking. For any queries and some of his transactions (acts), he visits the branch of his bank in the city (not his local one - scene), which has an International Trade Centre, and he knows a number of people there quite well (agency). He trusts them (act-agency). This trust has extended to online systems as the bank has contacted him when some anomalous activity occurred in his account (act-scene-agency).

Mark hates going into branches (scenes) and went on the Internet (scene) as soon as he could but believes security could be improved (agency). Nancy, likewise, took to IB as it liberated her (act-agency). She now works on her accounts (acts) while her children sleep or on Sunday afternoons (scenes). Her job running a small business became easier with IB (act-agency-purpose). Shane is another avid user of both the Internet and IB (scene-act). He recognizes the risk of IB but says any type of transaction is a risk (agency). Greg also follows his transactions (acts) everyday via IB (scene), although he will not use shared computers to access his account (scene-act). Samuel also conducts much of his business online but believes IB is not user friendly with too many clicks to confirm transactions particularly 'when you have forty of them' (act-agency-scene). Rita, on the other hand, is a reluctant user of Internet banking (act-agency-scene). Rita's husband has an SME but Rita is the one to deal with the bank. She still used an old passbook, which she had to forfeit, "I'm a traditionalist you see, I had to give it up actually" (act-agency).

From the preceding discussion, it becomes apparent that the group of SMEs in this study frequently speaks of the agency (that is, cheques, IB, cash) together with the scene (online, in the bank, at home on a Sunday afternoon). The acts they undertake (maintaining contact – saying 'hello', developing trustful relationships, depositing, keeping up-to-date with transactions) display an awareness of where they are being done (scene).

SMEs in Australia appear to have a love-hate relationship with their banks (agency-scene). Many of them take up IB (scene) but still visit the bank (scene). The scene of their banking thus becomes multiple with many of them maintaining the face-to-face contact despite not needing to. As shown by the variety of ways of interacting, some SMEs take advantage of the extended hours available through IB thus extending their scene (scene being both place and time dependent). The tendency of many to interact within a branch and to maintain their agency through this attendance places emphasis on the need to consider the scene for SMEs.

IB seems to increase the scope of agency allowing multiple forms although some are being denied to the 'traditionalist' (Rita with her passbook) and individuals who still use cash (Fred and Gay). SMEs appear to use IB both reluctantly and with fervour while still maintaining individually preferred forms of agency wherever possible. It is interesting to note that even IT-based businesses (6) in this study use a range of banking 'scenes'. Internet banking was only of the options used to do every day banking in an SME; other options regularly used were cheques and face to face banking.

6 Conclusion

Elaborating the pentadic elements with respect to banking allows the analysis to be focused on the multiplicity of parts of the communicative act that is banking. The scene is variously located, with the act taking place within a bank branch, outside a bank branch (ATM), online via a web interface or by phone. The scene in this case refers to the place where SMEs are interacting with their bank. The scene is expanding so that the bank branch is no longer the primary place of transaction for an SME. The agency of the individual within the SME incorporates the technologies: the

computer, the ATM, the telephone or the sociotechnical: the bank teller in face-to-face interactions. The majority of these SMEs are using IB and are feeling 'liberated' with its availability. This demonstrates that, unlike the findings of [6], who found online banking was virtually non-existent; a distinctly different cultural group of SMEs is emerging within the Australian context.

Banking is very personal, as is running an SME. Even though the literature from many countries around SMEs, such as Finland [11], Thailand [9] and South Africa [4] suggests that SMEs are hesitant and thus slow to adopt new technology, Internet banking may be an exception. It is being largely embraced and the reasoning suggested by our data is because of the personal benefits. The major benefits of Internet banking for SMEs seem to draw on time and economies - allowing more flexible banking practices, at any hour of the day.

This study also demonstrates that Burke's [5] Pentad provides a useful framework for discussing and describing a banking environment and for elaborating differences between cultural groups. The pentadic elements aim to ensure that all aspects are viewed. This means that attention is paid to the agent – the individual undertaking the act; the act itself; the method by which the act is undertaken (agency); the purpose for the act; and the location or scene of the act. Using Burke's terms provides a greater definition of who, where, when, why and how and allows the elaboration of interactions between the elements. In this way, it becomes clearer that the act, whether learning or interacting, cannot take place without the location or scene being included. The power of multiple perspectives, such as found in dyads (scene-act, scene-agent) is an explicit statement of the concerns that centre on a consideration of banking mediated by the interface of an online environment. This consideration proceeds from the assumption that no single perspective can provide the kind of analysis required to begin to comprehend the interactions of an individual with an Internet bank. This allows direct investigation of cultural factors.

Acknowledgments. Thanks to the Smart Internet Technology Cooperative Research Centre, particularly Supriya Singh and Liisa von Hellens, for supporting this research, and to the participants who willingly gave up their time and ideas.

References

1. Ågerfalk, P., Eriksson, O.: Usability in social action: reinterpreting effectiveness efficiency and satisfaction. In: Ciborra, C.U., Mercurio, R., de Marco, M., Martinez, M., Carignani, A. (eds.) Proceedings of the Eleventh European Conference on Information Systems, Naples, Italy (2003)
2. Beekhuyzen, J., von Hellens, L.: An actor-network theory perspective of online banking in Australia. American Conference on Information Systems (AMCIS), Acapulco, Mexico, Association of Information Systems (2006)
3. Beekhuyzen, J., von Hellens, L., Morley, M., Nielsen, S.: Searching for a methodo-logy for smart Internet technology development. In: 12th International Conference on Information Systems Development Methods and Tools, Melbourne, Australia, Springer-Verlaag, Heidelberg (2003)

4. Brown, I., Buys, M.: A Cross-Cultural Investigation into Customer Satisfaction with Internet Banking Security. In: SAICSIT '05: Proceedings of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries: South, pp. 200–207 (2005)
5. Burke, K.: A Grammar of Motives. University of California Press, Berkeley (1969)
6. Egan, T., Clancy, S., O'Toole, T.: The Integration of E-Commerce Tools in the Business Processes of SMEs, Irish Journal of Management (January 2003)
7. Glaser, B.G., Strauss, a.A.L.: The discovery of grounded theory: Strategies for qualitative research. Chicago, Aldine (1967)
8. Habermas, J.: The Theory of communicative action. Polity Press, Cambridge (1984)
9. Jaruwachirathanakul, B., Fink, D.: Internet Banking Adoption Strategies for a Developing Country: The Case of Thailand. Internet Research 15(3), 295–311 (2005)
10. Mejias, R.J., Shepherd, M.M., Vogel, D.R., Lazaneo, L.: Consensus and perceived satisfaction levels: A cross-cultural comparison of FSS and Non-GSS outcomes within and between the United States and Mexico. Journal of Management Information Systems 13, 137–161 (1997)
11. Pikkarainen, T., Pikkarainen, K., Karjaluoto, H., Pahnila, S.: Consumer Acceptance of Online Banking: An Extension of the Technology Acceptance Model. Internet Research 14(3), 224–235 (2004)
12. Purcell, F., Toland, J.: E-Finance for Development: Global Trends, National Experience and SMEs. Electronic Journal on Information Systems in Developing Countries 11(6), 1–4 (2003)
13. Ruth, A.: Learning at the Screenface: A pentadic analysis of email discussion lists, unpublished doctoral thesis, Griffith University (2004) Available at Australian Digital Theses: http://www4.gu.edu.au:8080/adt-root/public/adt-QGU20050316.170253/ index.html
14. Ruth, A.: The Screenface: Interfacial issues of mediating interaction Qualit2005 Challenges for Qualitative Research, 23-25 November, 2005, Institute for Integrated and Intelligent Systems, Brisbane (2005)
15. Singh, S., Beekhuyzen, J.: The Bank and I: Users' Perceptions of the Security of Internet Banking. Internet Research 7.0: Internet Convergences, Brisbane, Australia, Queensland University of Technology (2006)
16. Singh, S., Jackson, M., Beekhuyzen, J., Cabraal, A.: The Bank and I: Privacy, banking and lifestage. Computer Human Interaction (CHI) 06 Workshop on Privacy-Enhanced Personalization, Montreal, Canada, Institute for Software Research (2006)
17. Singh, S., Zic, J., Satchell, C., Bartolo, K.C., Snare, J., Fabre, J.: A Reflection on Translation Issues in User-Centred Design. 7th International Conference on Work with Computing Systems (WWCS 2004), Kuala Lumpur, Malaysia (2004)
18. Vredenburg, K., Isensee, S., Righi, C.: User-Centered Design: An Integrated Approach. Prentice Hall, Upper Saddle River, New Jersey (2002)
19. Weber, M.: Economy and society. University of California Press, Berkeley, CA (1978)

A Comparative Study of Thai and UK Older Web Users

Prush Sa-nga-ngam and Sri Kurniawan

School of Informatics, the University of Manchester PO Box 88, Manchester M60 1QD, UK
prush.sa-nga-ngam@postgrad.manchester.ac.uk,
s.kurniawan@manchester.ac.uk

Abstract. Numerous studies had pointed out the effect of culture on interactive system design and use. This paper reports on a study on the use and preference of web browsers by 100 respondents aged 50 years old and over from Thailand and UK, who arguably differ in their culture and online developmental curve. The questionnaire explored their online activities, browser manipulations, problems with standard browsers and features required. The study reveals differences in the types of activities these two groups of users performed online and in their preferences. The results of this study points to the need to design a culturally inclusive web browser in addition to an age-friendly web browser when dealing with older web browsers from different countries.

Keywords: web browser, ageing, questionnaire, culture.

1 Introduction and Background

Numerous studies had pointed out the effect of culture on interactive system design and use. For example, a study of cultural differences in the use of Instant Messaging in Asia and North America reported difference in usage and perception of IM e.g. audio-video chat, emoticons, and single vs. multi-party chat [1].

Older population is one of the fastest growing Internet users. According to Department of Economic and Social Affairs, United Nations, from 2000 to 2030, the world's elderly population (60 and older) will grow from 10% to 21%. Older people's adoption of the Internet also rose quite dramatically in the past decade. A survey conducted in February 2006 revealed that 72% of Americans aged 51-59 year-olds, 54% of 60-69 year olds, and 28% of 70-79 years olds went online [2]. So far, however, there has been little discussion about cultural differences in the older population's computer and Internet usage. Thailand and UK were chosen because both countries are diverse in (West/East) and in online development curve (Internet in Thailand and UK was introduced to the general public in late 1995 and 1991 respectively).

This paper seeks to investigate the cultural differences in the use of Internet and web browser by older adults in Thailand and UK.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 596–605, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 Stimuli and Participants

A questionnaire (in Thai and English) was designed to obtain information about the usage of Web browsers by older population. The sections of most interest to this paper are those focused older persons' usage pattern of their web browsers and problems faced by older persons when browsing the Internet.

The beginning of the questionnaire describes the questionnaire's aims, use of the data and instructions for completing the questionnaire. The main body of the questionnaire comprises both closed multiple choice and open-ended questions. It includes questions on demographics, internet and computer usage such as age group, gender, and duration of computer and Internet use. It also investigates the use of Web browser functions and user preferences. Space is provided at the end for comments on problems experienced when using Web browsers and the Internet, and further needs. The questionnaire was piloted with 2-3 respondents from both countries, after which minor revisions were made.

The questionnaire was printed on standard paper in black Tahoma 18pt. It was distributed during the months of March-April 2006. The requirements for participating are that they were 50 years old or older at the time of the survey.

The respondents consist of 53 Thais (44 Female/9 Male) and 47 UK (29 Female/18 Male). Out of the 53 Thai respondents, 43 were 50-54, 9 were 55-59 and one was 60-64 years old. A quarter of the UK respondents were 70-74, 9 were 60-64, 8 were 65-69 with the remaining four quarters spread equally in other age brackets. Table 1 provides the breakdown of their computer and Internet experience.

Table 1. Respondents' computer and Internet experience

	Thais		UK	
	No.	%	No.	%
Length of Computer use				
Less than 6 months	3	5.7	2	4.3
6-11 months	2	3.8	1	2.1
12-23 months	3	5.7	1	2.1
2-5 years	13	24.5	11	23.4
More than 5 years	32	60.4	32	68.1
Length of Internet use				
Less than 6 months	7	13.2	8	17.0
6-11 months	4	7.5	3	6.4
12-23 months	9	17.0	4	8.5
2-5 years	17	32.1	8	17.0
More than 5 years	16	30.2	24	51.1
Weekly Internet Usage				
Less than 5 hours	29	54.7	19	40.4
5-9 hours	10	18.9	6	12.8
10-19 hours	10	18.9	3	6.4
20 hours or more	4	7.5	19	40.4

3 Results

3.1 Internet Usage

When asked about the location for accessed the Internet (respondent could choose more than one locations, which are home, friend's or relative's computer, library or community centre, work and other location that they need to specify), 41 UK respondents checked home while 23 and 21 Thai respondent checked home and work respectively. Thirty-three (70%) and 16 (34%) UK respondents access Internet via broadband and dial-up respectively. Internet access of Thai respondents distributed closely among dial-up (34%), broadband (36%), and LAN (42%). This is an encouraging finding, as it indicates that older persons are using quite up-to-date connection technology.

One part of the Internet usage questionnaire investigated the purposes/topics for using the Internet. Some of the choices were derived from an article that reported the top 10 reasons of why older persons were online and an article from the Guardian newspapers that reported the online activities that older adults usually performed [3,4]. The most frequently chosen reason for using the Internet was to keep update with news and events. The least frequently chosen reason was to check stocks and investments. Table 2 provide breakdown by country and the whole sample. The most frequently reason for using the Internet of Thai and UK participants was to keep update with news and events and keep in touch with friends and family respectively. The Wilcoxon analysis revealed significance difference ($p<0.05$) in most topics except hobbies/interests, health information and stocks/ investments.

Table 2. Reason for going online. Number show Mean (S.D.) Options: 1 = everyday, 2 = twice a week or more, 3 = once a week, 4 = once every 2-3 week and 5 = once a month or less (or never).

	Thais	UK	Total	P
Business	2.77 (1.53)	3.70 (1.64)	3.21 (1.64)	.003
Stay in touch	3.68 (1.52)	2.32 (1.51)	3.04 (1.65)	<.001
News/Events	2.26 (1.36)	3.17 (1.65)	2.69 (1.56)	.007
Hobbies/Interests	3.15 (1.52)	2.89 (1.61)	3.03 (1.56)	.363
Health information	3.98 (1.28)	4.06 (1.26)	4.02 (1.26)	.772
Online shopping	4.89 (0.58)	3.91 (1.33)	4.43 (1.11)	<.001
Products/services	4.42 (1.12)	3.94 (1.15)	4.19 (1.15)	.005
Stocks/investments	4.79 (0.69)	4.43 (1.23)	4.62 (0.99)	.094

3.2 Browser, Browsing Devices and Windows

Expectedly as Microsoft Internet Explorer (IE) comes standard with Windows operating system (OS), all of Thai respondents (who know what their browser was

called) and 70% of UK respondents used IE as their browser. Only four respondents (3 Thais/1 UK) did not know what their browser was called. This indicates that close to 30% UK respondents went into extra length to install another browser (or use a different OS).

Significant difference in browsing device was found for the input device the respondents usually used to manipulate their browser (p=.035). More than half (58.5%) of Thai respondents used combination of mouse and keyboard to manipulated their browser compared with 38.3% of UK respondents. The percentage of Thai and UK respondents using mouse to manipulate their browser are 39.6% and 61.7% respectively. Only one Thai respondent reported to have used only keyboard for web browsing.

A question was asked about the number of browser window opened at one time, 58.5% and 36.2% of Thai and UK respondents said 2-3 windows; 22.6% and 38.3% of Thai and UK respondents said only one, the rest opened 4 or more windows.

The way for browsing long webpage is quite similar in both groups. Around half of both Thai (48.9%) and UK (50.9%) browsed long pages using the wheel of the mouse. 30.2% and 25.5% of Thai and UK respondents dragged scroll bar to browse long pages. The rest either clicked scroll bar or used Page Up/Down buttons. This indicates that the majority of older persons prefer controllable smooth and continuous page transition rather than fast and a page long transition.

3.3 Browsing Tasks

To investigate the functions in standard browsers that older persons used, the respondents were asked about 27 activities drawn by two HCI experts, who performed cognitive walkthrough of commercial web browsers (both Microsoft IE and Mozilla Firefox). For each activity, we asked whether the respondents had performed it (indicating weather it was performed with a mouse, a keyboard or both in combination) or not. Most activities were performed either using mice or combination with mice and keyboards. The most five functions found frequently used and unused are shown in Table 3.

The most frequently used functions are basic functions in web browsing. The functions used by the two groups are quite similar, except organizing favourite list or bookmarking for Thai respondents and stopping and reloading a webpage for UK respondents. The three similar unused functions relate to advanced functions and setting. The two different ones are setting browser's home and learning from browser's help or tutorial for Thai respondents and changing display language preference and changing text size for UK respondents. There is no significant differences in these 27 activities (Wilcoxon, $p<0.05$) except for six activities: open new browser window (p=.036), print web pages (p=.014), preview web pages before printing (p=0.25), Go back to previous page (p=.034), go to browser's default web page (Home page) (p=.042), and change display language preference (p=.022). The difference of the first five functions cause by a group of Thai respondents used those functions using a combination of mouse and keyboard while most of UK respondents use those functions using mouse only.

Table 3. Top used and unused functions

Thailand	UK
Top used functions	
1. Close web browser	1. Open new browser window
2. Open new browser window	2. Go back to previous page
3. Organize your Favourite or Bookmarks list	3. Go to browser's default web page
4. Go back to previous page	4. Close web browser.
5. Go to browser's default web page	*5. Stop and reload a webpage*
Top unused functions	
1. Set browser's advanced options e.g. set Java, ActiveX control	1. Set proxy server
2. Set proxy server	*2. Change display language preference*
3. View HTML source	3. Set browser's advanced options e.g. set Java, ActiveX control
4. Set your browser's home page	4. View HTML source
5. Learn from browser's help or tutorial	*5. Change text size*

3.4 Users' Mental Models

This part of questionnaire aimed at understanding users' mental model of various components of a webpage/website. In response to the question what gave away which website they were browsing, the most chosen object is the name shown on address bar and the second is a URL shown on address bar. Figure 1 shows country break down of the objects to identify a website and a link. Looking into difference by country, 32% and 47% of Thai and UK respondents chose the name shown on title bar, 36% and 15% of Thai and UK respondents chose a URL shown on address bar. The rest chose logo/banner, contents, and others.

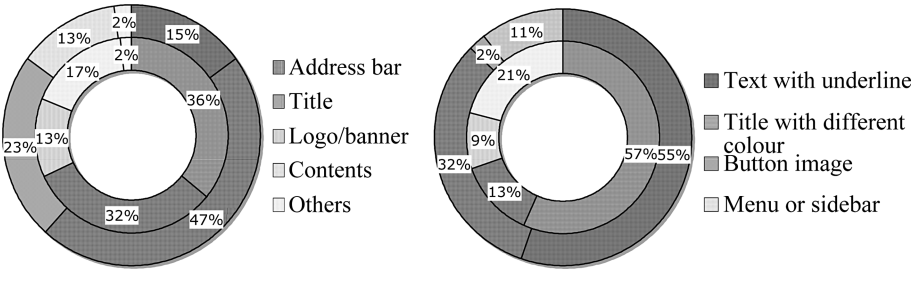


Fig. 1. Entity to identify website and link object (Thai=inner ring, UK=outer ring)

When asked about the object that gave away that an object was a link, more than half of respondent from both countries chose text with underline (Thai 57%, UK 55%), the rest chose text with different colours, button image and text or images in dropdown menu or sidebar. Mental model of users on the page loading status was asked through questions on whether the browser's status bar or the browser's animated logo provides useful information. Most of respondents (Thai 72%, UK 72%)

stated that the browser's status bar did provide useful information and (Thai 81%, UK 49%) said that the browser's animated logo provides useful information.

3.5 Problem and Difficulties

In response to the open-end question on problems and difficulties, we received some descriptions of problem and difficulties with web browsing shown in Table 4. As this question was optional, only 25 respondents offered their opinions that were categorized into 4 groups through content analysis. Most problems received from both Thai and UK respondents are related to website design and undesired content that respondents received. The different problems of Thai respondents were related to connection problem e.g. slow speed and cannot connect.

Table 4. Topics related to browsing problems

Thai respondents	UK respondents
Website and design (4) • Too much animation and text • Too much text and information. • Too much images and text • text display incorrectly **Undesired content (5)** • marketing, spam, promotion emails • Too much adds • Ads and pop-up • pop-up windows • ads pop-up **Connection (3)** • slow download speed • can not connect • slow speed of internet connection **Other (2)** • Update Version • Get viruses from internet	**Website and design (4)** • poor web design you come across while surfing. • Freeze Page not available 404 • This page is not available. • Some websites only work with Microsoft browsers. **Undesired content (6)** • aggressive marketing of annoying stuff - like pop-ups only worse • other than browser being slow at times, nothing. • Too Much spam. • All the Adverts blocking any progress • pop ups • Objectionable Content **Other (1)** • Using Password

3.6 Further Needs

To understand older persons' opinions on some features to assist their browsing more effectively, the respondents were asked to rate, in 5-Point Likert-like scales, from 'Must have' to 'Not needed'.

Figure 2 illustrates the distribution of ratings and comparing Thai and UK respondents. The most positively responded feature was the Pop-up window block with more than 70% rated must have or should have and the second was ads block. The most negatively responded feature was the Reminder (with 32.7% rated do not really need or not needed). The Wilcoxon analysis reveal significant different in most

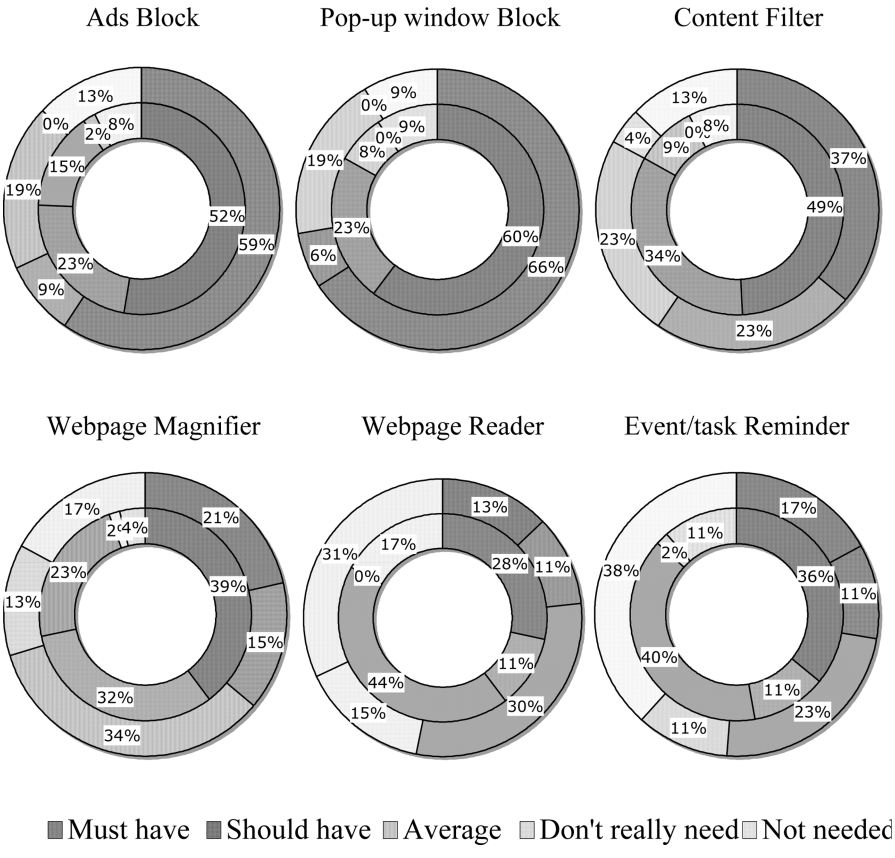


Fig. 2. Respondents' opinions on browser features (Thai=inner ring, UK=outer ring)

feature except Ads block (p=.906) and Pop-up window block (p=0.876). The need of webpage magnifier is highly significant different (p<.001). Upon closer inspection 71% of Thai respondent required webpage magnifier while only 36% of UK respondents require this feature.

4 Discussion and Conclusions

This paper seeks to investigate cultural differences in the use of Internet and web browser by older adults in Thailand and UK.

4.1 Reason for Internet Usage

The findings on Internet usage to keep in touch with friends and family and online shopping show highly significant difference (p<.001). The difference in the fist reason is due to cultural difference. Thai families and friends (and in general Asian families and friends) usually live in the same area. Thus, the need to communicate via the Internet is less pronounced than in Western culture. Using the Internet for online

shopping is very unpopular in Thai respondents (5.7%) compared to the case of UK respondents (31.9%). This may be because Thai people are still unfamiliar with online shopping due to issues of trust, safety and availability of online store and Internet access. This finding is similar to a Thai Internet user survey in 2005 [5].

Using Internet for hobbies/interests, health information, and stocks/investments are not significantly difference. Those tasks are common reasons for Internet use of older adults in both countries. Checking stocks and investments is similarly unpopular in both groups, possibly because older people have concern about security and are afraid to make errors. In addition, stock trading in not popular in Thailand in general.

4.2 Browser, Browsing Device and Window

Nielsen//NetRatings report on the share of UK browser's market (2006) reported that IE is the main browser for 88% of UK users [6]. The result is somewhat echoed in this study (70% of UK respondents using IE). The Thai internet survey reported 93% of Thai users use IE [5]. The result is much echoed in this study when considering all of Thai respondents using IE.

It is unexpected finding that 58.5% of Thai respondents mainly use a combination of mouse and keyboard during web browsing while 61.7% of UK respondents use only mouse to manipulate their browsers. One possible reason is that the browser interface language is English, which might not be familiar to some users. Remembering and using keyboard shortcuts might be easier for this user group. Further study is required to verify and understand exact reasons.

The findings show that 58.5% and 36.2% of Thai and UK respondents open 2-3 browser windows at one time. This finding contradicts the guideline suggested by Kurniawan and Zaphiris to provide one window for older web user [7]. However, higher percentage of Thai respondents that open 2-3 browser windows might be caused by slower Internet speed. Users may open the second and the third window to allow more time for webpages to load while working with the first window.

4.3 Browsing Tasks

Expectedly, most frequently used function of both Thai and UK respondents are basic functions required during web browsing. The only difference found is the organization of favourite or bookmark list in Thai respondents and the stop and reload webpages in UK respondents. Again, one possible cause is the difference in Internet connection speed. Slower Internet speed in Thailand might cause users to rely on using their bookmarks while UK respondents can start browsing from their favourite search engine like Google or Yahoo!

The three similar unused functions are related to advanced functions and setting. The two different functions are set browser's homepage and learn from browser's help or tutorial for Thai respondents and change display language preference and change text size for UK respondents. Thai user do not set browser's home page because of the same reason we mentioned previously, they mainly rely on their bookmark, so they rarely or never set the browser's homepage. All other unused functions are related to language difference, browser's help are by default presented in English, which renders it unusable for some Thai users. UK users do not change

display language because they may not require to browse webpage in other languages or the browser can change the language preference automatically when browsing webpages in other languages. Changing text size might rarely be required because Latin alphabets are easier to read in their default setting than Asian characters like Chinese, Japanese including Thai are.

4.4 Problems and Difficulties

Problems faced by Thai and UK respondents are quite similar in term of website and its design e.g. page clutter, poor web design and undesired contents e.g. spam and ads pop-ups. The only difference is that some Thai respondents also have connection problem in term of speed and reliability. Software developer might consider special feature to address these problems e.g. giving clearly visible or audible feedback during webpage loading and alert users when loading is done, and alert them when connection problem occurs.

4.5 Further Needs

Pop-up window and ads blocks are clear winners in term of further needs. It indicates that undesired content screeners featured highly, much more than visual aids. It is interesting that 71% of Thai respondents required webpage magnifier compared to 36% of UK respondents, even though the Thai respondents are younger than the UK respondents. One possible reason is because Thai language has the anomaly of not using spaces to segment syntactic units and spaces are used only to delimit sentences. Figure 3 shows an example of Thai text. Thus, small Thai text can be problematic to older people, even after their vision is corrected with glasses.

ภายหลังการเดินทางไปตรวจเยี่ยมโรงงานสร้างดาวเทียม "ธีออส" ดาวเทียมสำรวจทรัพยากรธรรมชาติดวงแรกของไทย ณ ประเทศฝรั่งเศส เมื่อสัปดาห์ที่ผ่านมา และการตรวจเยี่ยมศูนย์ดาวเทียมภาคพื้นดิน สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ เมื่อวันที่ 15 ม.ค.ศ.ดร.ยงยุทธ ยุทธวงศ์ รัฐมนตรีว่าการกระทรวงวิทยาศาสตร์และเทคโนโลยี เปิดเผยว่าขณะนี้ดาวเทียมธีออสได้สร้างแล้วเสร็จ และกำลังอยู่ระหว่าง

Fig. 3. An example of Thai text

4.6 Summary

In summary, there are cultural differences in the use of Internet and web browser by older adults in Thailand and UK. And as many studies comparing two populations often suggested, the safest bet for developers of systems that would be used by these two populations is to take the lowest denominations. In this study, it means ensuring that all of the further needs are addressed, all of the problems experienced by the two groups rectified, and the simplest design that would allow loading at low connection speed is facilitated. This is not an exhaustive list, just a starting point to a culturally-inclusive design.

5 Limitations of the Study

There are naturally some limitations of this study, mostly related to sample's demographics. The Thai respondents are relatively younger than UK respondents. However, both group of respondents' computer usage and Internet experience are quite similar. This should compensate the age limitation at a certain degree.

The gender split is rather typical of voluntary studies of older adults, with more women than men participating, but would need addressing in further studies.

Using questionnaire for data collections caused difficulty in flexibility of the questions. The questionnaire still lists rather limited set of feature and activities that we could investigate, even we consulted and expert older web user and run a pilot study. Another inquiry method such as focus group discussions or interviews would complement the data collected from this questionnaire very well.

References

1. Kayan, S., Fussell, R.S., Setlock, D.L.: Cultural differences in the use of instant messaging in Asia and North America. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 525–528. ACM Press, New York (2006)
2. Fox, S.: Are "Wired Seniors" Sitting Ducks? Pew Internet memo (2006) from http://www.pewinternet.org/pdfs/PIP_Wired_Senior_2006_Memo.pdf
3. Reaves, K.: Top 10 Reasons Older Adults Should Be Online (2006) from http://www.family. org/focusoverfifty/articles/a0020568.cfm
4. The Guardian: On the crest of a wave. Technology section (2004) from http://technology.guardian.co.uk/online/story/0,3605,1353207,00.html
5. National Electronics and Computer Technology Center. Internet User Profile of Thailand 2005. National Electronics and Computer Technology Center, Bangkok (2006)
6. Nielsen//NetRatings: Battle of the browsers: IE VS Firefox. (2007) from http://www.netratings.com/pr/PR_122006_1_UK.pdf
7. Kurniawan, S.H., Zaphiris, P.: Research-derived web Design Guidelines for Older people. In: Proceedings of SIGACCESS 7th International Conferences on Computers and Accessibility (ASSETS'05), ACM Press, New York (2005)

A Qualitative Oriented Study About IT Procurement Processes: Comparison of 4 European Countries

Michael Schiessl and Sabrina Duda

eye square GmbH, Schlesische Str. 29-30, D-10997 Berlin, Germany
schiessl@eye-square.com, duda@eye-square.com

Abstract. This study shows that in a qualitative study a small sample size is sufficient to gain interesting results and show differences between the procurement of IT services in different countries. It demonstrates how a requirements analysis can be conducted in a very early phase of a web development project. For the website of an IT company which sells products worldwide the needs of potential users of its website should be identifed. With 8 users each from Germany, Switzerland, Belgium and Spain, an in-depth interview was conducted; after that some page drafts were shown. The users were from different hierarchy levels and from companies of varying sizes; all involved in the IT procurement process. The study showed who is involved in the different stages and what is relevant in each stage. The distribution of responsibilities was different in each country. The study gave insights into how to support potential buyers during the IT procurement process and how to adapt the web pages to local needs.

Keywords: International Usability, Requirement Analysis, IT Procurement, Web Usability, User Test.

1 Introduction

An international IT company with headquarters in an English speaking country has conducted a usability study in different European countries during the redesign of their website. The study was conducted at a very early stage of development. The website shall support employees of companies who plan to buy IT systems. A theoretical model about the processes of IT procurement was already in existence. However it was unclear to which extent employees are involved in each of the six stages; furthermore, it was unclear how tasks were split. This study should shed light on how the real IT procurement processes are carried out, with a sample of employees from different hierarchy levels (from Technician to CTO). This study focused on the results of in depth-interviews. Additionally some already existing drafts of the web pages were evaluated; here it was especially important how users found their way from the company page to the country specific page.

2 Research Questions

For the website of an IT enterprise which is selling its products worldwide, the needs of potential users of the website were to be identified.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 606–614, 2007.
© Springer-Verlag Berlin Heidelberg 2007

- How is the process of IT procurement going on? Who is involved in which stages of IT procurement?
- What is important when choosing a supplier?
- How do users like the example web pages?

3 Method

With eight users each in Germany, Switzerland, Belgium and Spain an in-depth interview (lasting 2 hours) was conducted. These users were involved in the IT procurement processes in their company. The users were from different hierarchy levels, and the company size was varying. In total we had 32 subjects (31 male, 1 female): 2 x CEO/CTO, 9 x Senior IT Manager, 8 x Operational IT Manager, 8 x Technical Specialist, 5 x Project Manager. Mean age was 36 years (25 – 61). All had received a higher education.

These are the stages of IT procurement we explored in detail:

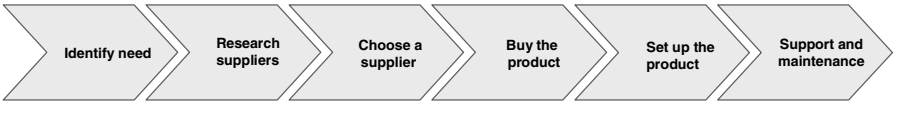


Fig. 1. Stages of IT Procurement Process

4 Results

4.1 Stages of IT Procurement – Who Is Involved?

Results surprisingly showed that only the very highest level and the lowest level of the hierarchy are involved in only a few stages: The few top-level subjects (CEOs) of the sample are undoubtedly more involved in the earlier stages; low-level positions (e.g. technicians) are involved in fewer stages (research, set-up and maintenance).

The higher levels of the hierarchy are involved in almost every stage of IT procurement; especially in Germany and Switzerland, where the higher management seems to have a need to control all stages.

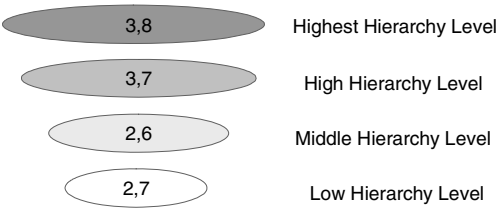


Fig. 2. Mean number of stages in which employees are involved: All countries

In Germany and Switzerland the subjects tended to be involved in more stages. The higher up the ladder they are, the higher their involvement is! The people with higher

responsibility in Germany and partly in Switzerland care about everything. They may not delegate as much as those in the same position in Spain and Belgium. In Belgium and Spain people seem to be concentrated more on their specific tasks. Highest management in Germany wants information about every stage of the process. The Germans have a high demand for technical details. They are more concerned about controlling everything.

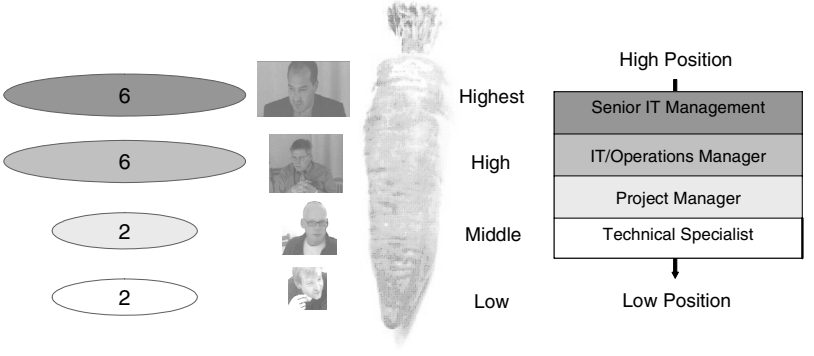


Fig. 3. Mean number of stages in which employees are involved: "Carrot Model" Germany

In Spain and Belgium, the management wants only the information they really need. The image is the most important information for the CEOs. They are less interested in details. In Spain the people responsible have stricter divisions concerning their tasks! High management only has a few tasks; usually to identify needs.

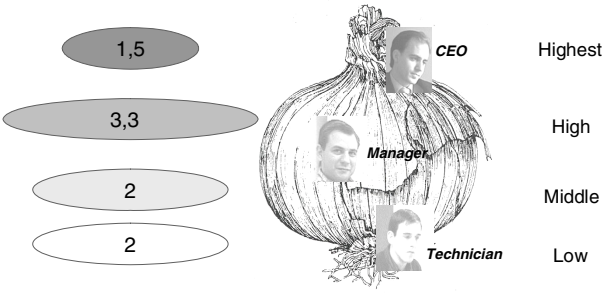


Fig. 4. Mean number of stages in which employees are involved: "Onion Model" Spain

The identification of the need is mostly initiated by interior triggers. External facilitators like marketing material and websites play a very minor role.

In the first stage of procurement: identifying needs, people from all levels in the hierarchy are involved (the management is always involved). The lower levels (technicians) tend to be involved in the last stages such as, set up and maintenance.

	Highest Level	High Level	Middle Level	Low Level
Germany	6	6	2	2
Spain	1,5	3,3	2	2
Belgium	3	2,5	2,5	3
Switzerland	4,5	3	4	4

Fig. 5. Mean number of stages in which employees are involved: All countries

4.2 Identify Needs

Needs are typically identified through a customer demand, technical reasons or for the purpose of cost cutting. The impulse for the idea for a new system is not generated externally (e.g. marketing, vendors), but internally (e.g. professional department, co-workers, clients, ideas of the management).

Usually everybody is involved because it is an important issue (the management is always involved). This reveals that the marketing should be directed at all levels in a hierarchy (conveyance both of image and technique).

4.3 Research Suppliers

The search for a supplier is usually initiated either through contacting a known supplier - via phone preferably! - or one that has been recommended (many have established relationships, fixed suppliers).

Quickly this list is reduced by taking in technical and pricing considerations. At this point, the web is sometimes used to check references or company sizes, but less often to check products. The subjects do not usually look for products on the supplier's website, but rather check other websites (for field reports, case studies, the financial standing of the supplier) because they are neutral.

During the phase of searching for a product, presentations of the product or visits to other clients with the product already up and running are important. On average a rather small amount of 3-6 suppliers were actually contacted. The others were quickly eliminated based on reasons like inappropriate techniques or too high prices. The first contact with a supplier must be recognized to be competent, offering personal presentations of products or on site inspection.

Quotes

- "Normally we work with 3 suppliers because we have the tendency to trust more the 'familiar' suppliers.. who already know us, […]. It is important to work with several suppliers to generate competitiveness between each one of them." (Spanish user)
- "[…] first of all you have to look at the company references. Then you have to find out if it is a small business (eg.20 employees) or a big business (eg. 2000 employees). In today's economic situation there are too many companies being

bought out and as a result products are no longer supported. The point is, I need a company with a certain market presence and reference. Surely we place orders between 5-10 thousand Euros; when it is more we really look for these things." (Geramn user)

4.4 Choose a Supplier

When deciding on a supplier, technical matching, flexibility, adaptation of the product to the individual needs, service and service contracts very deemed priorities and only at the very end the is the price considered important. Top decision makers expect to be contacted and get special individual service. They do not want to look through marketing material or search the web.

The information that is important when choosing a product is similar to that of when searching for a product. The technical matching of the product (does it meet the desired requirements), whether the price can be individually customized (adaptation), and the validity of the servicing and warranty must be explored. Service and flexibility are absolutely crucial when actually choosing a supplier. When the client has chosen a supplier sometimes a phase of negotiation starts (extras, individual adaptations, service contracts).

Important factors when choosing a supplier are whether a good service is offered in combination with flexibility (of products). This dictates that it should be possible to buy separate modules (not only total packages). The system should be able to be integrated into existing parts (compatibility) and should be able to adapt to a client's individual needs.

To govern this efficiently should assign one steady key-account manager from the beginning onwards. Investment in technically well-trained account managers embodied with the authority to make an individual pricing policy appears to be promising. For higher management a phone call or personal presentation would be convincing.

Because higher management have little time and do not bother to search actively they like to have things presented to them. They like to get specific, tailored and condensed information via phone (quicker than the web or studying marketing material). Technical specialists and lower-level positioned people will be convinced more by technical facts, case studies, test reports.

All in all, former state-run telecom companies and big international companies (e.g. IBM, Microsoft) are considered as being relevant suppliers. Other companies have to be present locally and offer service on site. It is important for international suppliers to provide service on site. They should have a local branch and local contact partners. People worry about availability of quick service on site and spare parts in their country.

In the study it was discovered that what seemed to hinder most people in seeing the IT company of this study as a relevant supplier is the fact that they seem to have had relatively little exposure to it. In Spain and Belgium everyone knew them and in Switzerland 2/3 of the subjects, but in Germany only half of the subjects knew the company.

Quote

- "Well, I think the bottom line is that they are not very well spread out in Germany. They are not present in Germany or Berlin. They don‘t offer the same service that they do in their home country." (German user)

4.5 Buy the Product

The management and the financial department confirm their decision - then the product is bought. Most common is phoning (oral confirmation of order) and faxing the signed offer. Sometimes a personal meeting is arranged for signing a contract.

Usually there is no particular procedure taking place immediately after the purchase is confirmed. The client waits for the product and expects the supplier to deliver on time. Occasionally further price negotiations take place. The transparency of the delivery status is important, and the client must be informed early if the timeline cannot be kept.

During the buying phase if everything is going well, there is practically no communication between the supplier and client.

4.6 Set Up the Product

During set-up in almost all cases the supplier is on site and installs the product. The reason behind this is that they have to guarantee the product and therefore prefer to do it by themselves. During this phase the contact is mostly face-to-face (sometimes via phone if supplier is not on site). It is important that there is one technician who remains the responsible contact person allowing familiarity and consistency for the client.

4.7 Support and Maintenance

When referring to maintenance the personal key-account manager is most important and should ensure that the service should be quick, reliable and around the clock (24 h/day and 365 days/year). The service contracts are very important and are vital for companies to consider using a supplier again. Support is so crucial that a lack of it can be the starting point for considering another supplier.

Online tools are rarely used and the use of an online tool is not seen. An online tool is too much of an effort especially for people higher in the hierarchy.

Quotes

- "I want to know how fast they can react in an emergency and replace parts and what parts of Germany they cover. Even if they are a small supplier could they at least do the job within 6 or 7 hours? Or do they have to bring the replacement parts from abroad or is there a warehouse here?" (German user)
- "...maybe it has something to do with the fact that at Camcom there is a contact person who I can call when something doesn‘t work. When you are dealing with

5000 Euros, 100 Euros more or less really doesn‘t matter. What is more important is that there is someone there to phone than to be really lost or desperate." (German user)

4.8 Evaluation of the WebPages

Reactions to the example web pages were mixed: In Germany users were more negative about the homepage; in Spain users were more positive; in Belgium and Switzerland comments were more balanced. Few users understood how to get to the specific country site; many complained that the path to the products is too long.

The website should contain country specific content and language (as well as country specific URL; e.g. ".de"). Country specific language is more user-orientated and language is an indicator for country specific content. Websites are not generally used by top managers, people in high positions seem to not do very much research on their own on websites; they might delegate that to people lower in the hierarchy who are not as fluent in English. There must be country (not language!) selection on homepage. Only those products and services should be on the homepage which are available worldwide or hint that this is only an overview and user has to go to his country site to check availability.

Quotes

- "We like to operate with suppliers that are nearer to us and know about us and our problems, language is important in facilitating confidence." (Spanish user)
- "Like I said, since the end user knows his own language the best, the selection should be at the beginning and should be in German. Maybe there should be a portal introduced and from there it should branch out into the different languages." (German user)
- "'Choose your country/language' must be presented on an earlier page (home) in order to avoid confusion. As a Swiss I am interested in Swiss products, therefore the choices have to be narrowed down right from the beginning…" (Swiss user)
- "I would prefer the “Select your country” in the combo box[...] Right now I think I will get the same page translated into Spanish, but not necessarily a specific page for Spain.” (Spanish user)
- "What confused me was, are the products and services on xy.com also available in Switzerland or not?" (Swiss user)

The link "Global Products and Solutions" which leads to the country-specific pages is unclear to the users. Almost every subject misinterpreted the meaning of this link (they were misled by the word "global"). The supplier determined everything outside its home base as global; therefore the link on the corporate page is named "Global Products and Solutions". But for the people in Germany, Spain, Belgium, Switzerland their country is not global, but local. Almost every subject misinterpreted the meaning of this link (they were misled by the word "global").



Fig. 6. Misinterpreted link on example web page

Quotes

- "To me, Global Products and Solutions means solutions for the whole company, for all departments." (Spanish user)
- "I honestly can't imagine what it means. "Global Services"…uh, world-wide data connection, etc..." (German user)
- "Probably solutions for companies with subsidiaries spread across the world." (Belgian user)
- "I think that xy offers a global solution for the whole company in a specific application." (Spanish user)

4.9 Branding

From these results it should be gathered that campaigns should focus on image and brand establishment. E.g. show a consultancy situation: satisfied clients talking with the company's account manager. Building on this and as equally important are brand recognition, great service and being local situated. They should be a local presence and it is here marketing and advertisements are important.

Quote

- "For me, I think its primarily quality. The people who put effort into their advertisements and get across their ideas psychologically, I think would probably also make quality products." (German user)

5 Summary and Recommendations

The Internet is not of primary importance during the procurement process. The suppliers site is useful for checking a company's size, references, case studies and information about products in the earlier stages (financial standing is - of course - checked on other sites) and later in the process for finding contact information.

During the early stages the subjects regard objective evaluations and test reports of the product on other unbiased websites as more reliable. Typical tasks when visiting a supplier's website (as identified in the study) were to find information about products and to find contact information.

Marketing material was found to be of minor importance; it did not help any of the subjects to initiate buying. However marketing must be intensified to enhance the

perceived presence of the company in each country and to raise local trust in the company and its products. In this respect marketing should be wary of two phases - for the first stages of the IT procurement process it is important to perceive a supplier as being local and on site. During the following stages objective facts about technical features, performance and reliability of the supplier's products gain more importance.

Marketing for the first stages can be concentrated on image. This will have an indirect effect on buying activity. When advertisements are present people will expect that the company branches and account managers are also present. This is considered the first step that will motivate people to add the supplier to their relevant set of suppliers.

The enhancement of the perceived local presence of the company can be achieved through marketing and advertising. The content of such should show account managers in consulting situations and show contact telephone numbers.

For the next stages - the real buying decision - objective, neutral reports are crucial (test reports with benchmarking, case studies, testimonials and experience reports from satisfied clients). In order to be globally successful, one has to be local. For example a local image crossed with a global idea (make the marketing witty: eg. show a British Guard saying "I think Bavarian"). E.g. successful Ikea campaign in Germany plays with being Swedish and at the same time being local: ‘Berlin is the Swedish’; Turkish people with blond wigs are shown.

An Empirical Study on the Smallest Comfortable Button/Icon Size on Touch Screen

Xianghong Sun[1], Tom Plocher[2], and Weina Qu[1]

[1] State Key Laboratory of Brain and Cognitive Science, Inst. of Psychology, Chinese Academy of Science, Beijing 100101, China
{sunxh,quwn}@ psych.ac.cn
[2] Honeywell ACS Labs, Minneapolis MN 55418, USA
tom.plocher@honeywell.com

Abstract. For the convenience of firefighters' decision-making and operation, touch screen display was chosen as the preferred interface for a fire information display system. Few studies were conducted to determine comfortable button/icon size on touch screens. This experiment investigated the effect of four factors on operators' performance with touch screen: 1) button size (20*20, 30*30, 40*40, and 50*50 pixels 2), spacing between buttons (0, 5, 10, and 20 pixels), 3) button/icon types (digit buttons only, picture icons only, combination), and 4) glove wearing (wearing vs. not wearing). 14 males were asked to accomplish a series of matching tasks on touch screen with the forefinger of right hand. Results showed that the spacing between buttons/icons, and wearing or not wearing a glove did not affect performance. Subjects pointed to the digit buttons faster than the other two kinds of buttons/icons. There was a significant difference among button/icon sizes. People performed best when it was equal to or bigger than 40*40 pixels.

Keywords: touch screen, button size.

1 Introduction

The touch screen is not used as an input device nearly as widely as the mouse. The mouse is very easy to use for dragging, drawing, and accurately pointing to some specific position on the screen [2]. Touch screens lack pointing accuracy, but are very intuitive for consumers [6]. They are especially useful for novices, who may be unfamiliar with mouse and keyboard operations, and for software systems with very limited control interactions [5]. User interface designs for touch screens must carefully consider the size of and spacing between touch-activated buttons and icons so that the user's inputs will be accurate. Usually, the larger the button, the easier it is for users to accurately point to it. But often, computer screen space is limited. Designs must trade off between button size and spacing that maximizes accuracy, and the ability to support the desired functionality for a given screen [2]. Those tradeoffs need to be guided by knowledge of how button size and spacing affects performance. Bender and Gregory [3] investigated people's performance on touch screen, they found an appropriate auditory feedback signal might help compensate for the reduced auditory feedback and increase touch screen performance. From other literature we found, few studies have been done that provide such design guidance [1].

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 615–621, 2007.
© Springer-Verlag Berlin Heidelberg 2007

In this experiment we investigated the smallest comfortable button size on a touch screen for firefighters' to activate using their forefinger. Four factors were considered:

1. The size of the button. With the resolution of 1280*1024, there were four levels: 20*20, 30*30, 40*40, 50*50 pixels.
2. The spacing between any two buttons/icons, there were also four levels: 0, 5, 10, and 20 pixels. We think there could be some trade off between the button size and the spacing between buttons.
3. The type of buttons/icons. Three levels: 10 digit buttons only, 10 picture icons only, and a combined case with the 10 digit buttons and 10 picture icons together. The ten digit buttons were: 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. The ten picture icons were: .
4. Wearing vs. not wearing a cotton waved glove. Firefighters usually wear some special uniform and a pair of gloves to the site. Typically, they have no time to take off their gloves before using the touch screen at the fire scene.

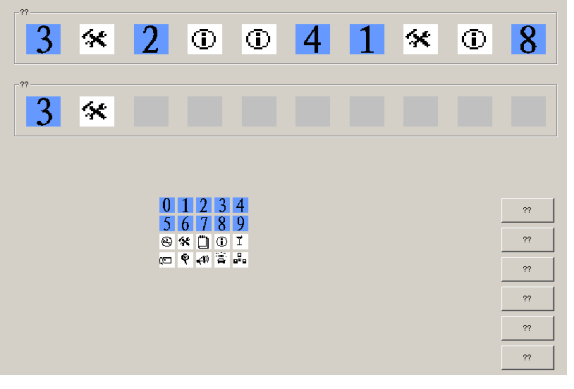


Fig. 1. Sample experimental task screen

2 Method

2.1 Experimental Task

On the computer screen, 10 buttons appeared in a stimulus row at the top of the screen. The buttons in this row consisted of 10 digit buttons or 10 picture icon buttons, or a combination of 5 digit and 5 picture icon buttons. The order of the 10 stimulus buttons was randomised from trial to trial. Below this, in the middle of the screen, appeared a keypad. The keypad consisted of the same 10 digit buttons in order from 0-9 and the same 10 picture icon buttons, arrayed in a 4 x 5 matrix. Subjects were asked to use the keypad to match each button in the stimulus string at the top of the screen, using their forefinger to point and select matching buttons from the keypad. The size and spacing of the buttons on the keypad was varied from trial to trial. The subject's selections from the keypad were displayed, one by one, in a results row immediately below the stimulus row. The reaction time to each button/icon and the percentage of correct inputs were recorded automatically by computer software.

Figure 1 is a sample screen from the experimental task and shows the stimulus row at the top, the results row below it, with two of the ten responses completed, and the 4 x 5 button keypad used to perform the matching task.

2.2 Experimental Design

The experiment used a 2 (wearing vs. not wearing glove)× 3 (digit buttons, icon buttons, mixed digit and icon buttons) × 4 (button sizes 20*20, 30*30, 40*40, 50*50) × 4 (button spacing 0, 5, 10, 20) within subjects design.

2.3 Participants

Since all the firefighters in China are male soldiers, and their age is around 18 to 36 years, 14 male subjects at the same age participated in this experiment. Each subject took half an hour to accomplish the experimental task.

Table 1. Percentage of correct inputting digits/icons with forefinger on touch screen

		Not wearing glove				Wearing glove			
Size	Inter-val	digit	icon	digit+ icon	total	digit	icon	digit+ icon	total
	0	83.8	86.0	88.5	86.1	74.3	72.7	77.5	74.6
20	5	96.4	95.3	94.6	95.5	91.3	93.3	94.5	92.9
*	10	92.9	96.4	97.9	95.7	91.3	95.6	91.8	93.1
20	20	95.0	88.3	95.0	93.0	93.8	97.9	95.6	95.8
	total	92.2	91.6	94.1	92.6	87.7	89.8	89.3	88.9
	0	96.9	93.3	98.3	96.0	90.7	89.4	92.5	90.7
30	5	97.9	96.7	93.1	96.0	96.4	95.3	99.1	96.8
*	10	90.8	98.7	92.9	94.3	95.8	98.0	95.5	96.6
30	20	91.4	99.3	90.8	93.9	95.7	93.8	96.4	95.1
	total	94.3	96.9	93.7	95.0	94.6	94.0	95.8	94.7
	0	98.6	98.6	96.9	98.0	97.7	95.6	99.2	97.3
40	5	99.2	99.3	97.1	98.5	97.9	98.8	94.5	97.3
*	10	97.5	98.0	99.2	98.3	94.3	95.0	96.7	95.2
40	20	93.1	98.6	92.9	94.9	98.7	95.6	100.0	97.9
	total	97.1	98.6	96.5	97.4	97.1	96.3	97.6	96.9
	0	93.6	99.3	97.7	96.8	94.7	97.5	96.7	96.3
50	5	99.2	98.7	100.0	99.3	99.3	99.3	99.1	99.3
*	10	96.7	98.7	90.7	95.4	96.9	96.3	100.0	97.5
50	20	93.1	96.4	92.9	94.1	99.2	98.7	100.0	99.3
	total	95.6	98.3	95.3	96.4	97.5	97.9	98.9	98.0
	0	93.3	94.1	95.3	94.2	89.3	89.0	91.5	89.8
total	5	98.1	97.5	96.3	97.3	96.1	96.7	96.8	96.5
	10	94.3	98.0	95.1	95.9	94.4	96.2	96.0	95.6
	20	93.1	95.9	93.0	94.0	96.9	96.4	98.1	97.0
	total	94.7	96.4	94.9	95.4	94.2	94.6	95.5	94.7

3 Results

3.1 Percentage of Correct Inputs

Table 1 shows the percentage of correct inputs for the digit/icon matching task. From the data, the percentage correct was stable whether or not the subjects wore gloves and no matter what kind of string, digit or iconic, the subjects were asked to match. But it increased significantly from 89% to 98% with the button/icon size ($F= 29.2$, $p = .000 < .01$) and with the button/icon spacing ($F= 14.1$, $p = .000< .01$). From the post hoc test, it was found that the percentage reached a stable point when button size was equal to, or bigger than 40*40 pixels. Compared to the other three spacing, only when the spacing size was zero was the percentage lower.

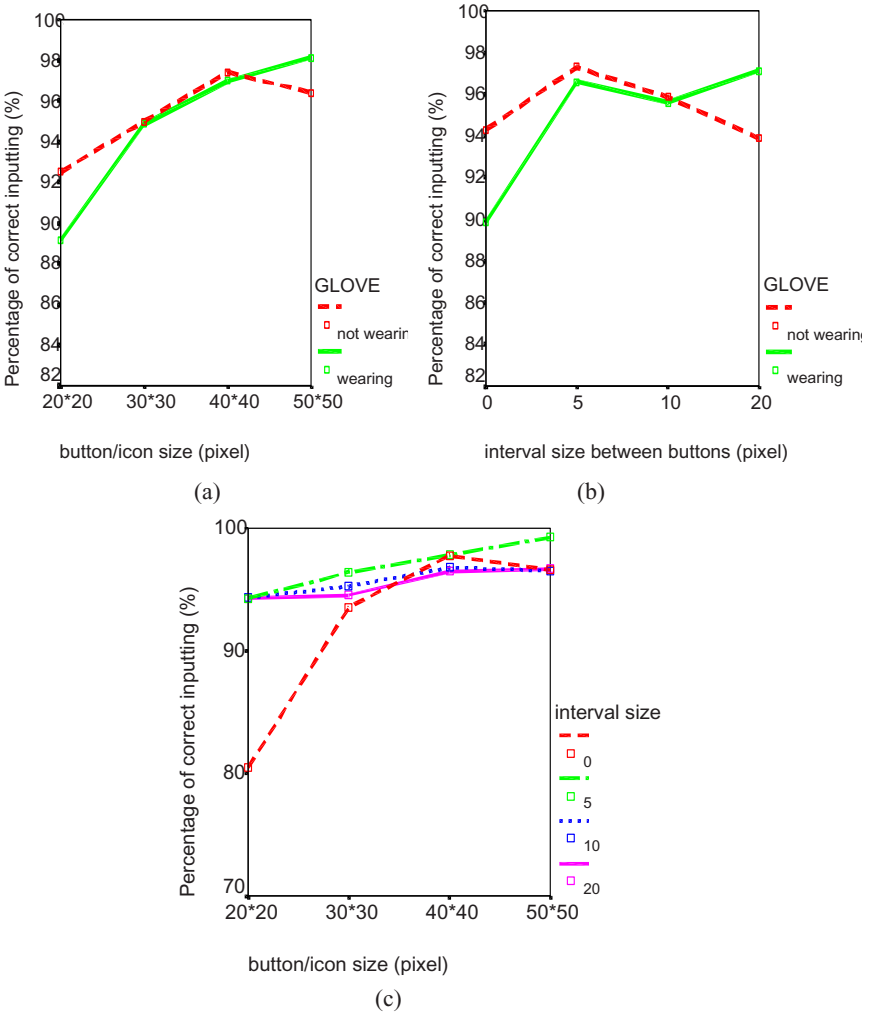


Fig. 2. (a-c) Variation and interactions in percentage of correct inputs between different factors

The interaction between button spacing and button/icon size was significant ($F = 9.1$, $p = .000 < .01$). Also the interactions between button spacing and glove-wearing ($F = 7.8$, $p = .000 < .01$) and between button/icon size and glove-wearing were significant ($F = 3.6$, $p = .014 < .05$). As shown in 2(a) and 2(b), the percentage of correct inputs was lowest at the smallest button size (20*20 pixels). Input performance improved differentially as button size and interval increased for subjects who wore a glove compared to those who did not. In Figure 2(c) it can be seen that the percentage correct also was lowest at the smallest button size, but increased differentially compared to other button spacing as button size increased. The percentage of correct inputting remained almost constant, when the spacing was 5, 10, and 20 pixels.

3.2 Reaction Time (RT)

Table 2 listed the mean reaction time to each button on the touch screen when subjects followed the stimulus string and matched them with the forefinger. All the reaction times ranged from 800ms to 1800ms. Based on the ANOVA for repeated measures, it was found that the RT was not affected by glove-wearing or by button spacing. However, RT was 200ms faster when matching digital stimuli as compared to iconic or mixed digital and iconic stimuli ($F = 59.0$, $p = .000 < .01$). The RT was shortened significantly as button/icon size increased ($F = 148.0$, $p = .000 < .01$). From the post hoc test, it was found RT to buttons at 30*30 pixels was significantly faster than to buttons at 20*20 pixels. RT to buttons at 40*40 and 50*50 pixels was significantly faster than to buttons at 30*30 pixels. RT was fastest when button size was equal to, or greater than 40*40 pixels. But considering the trade-off between the button size and cost of RT, the size 30*30 pixels was also a good choice for accurate and comfortable operations.

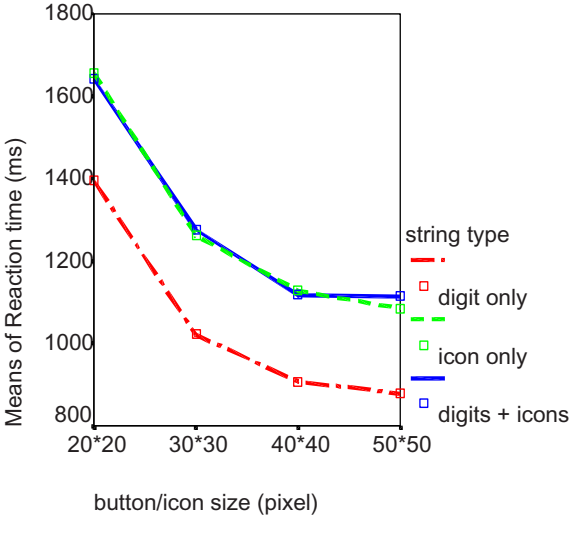


Fig. 3. Variation of RT with button/icon size increased when coping different types of strings

Table 2. Reaction time for inputs of digits/icons with forefinger on touch screen

Size	Interval	Not wearing glove			Wearing glove		
		Digit	icon	digit+icon	Digit	Icon	digit+icon
20*20	0	1325.5	1682.4	1708.8	1382.7	1665.5	1630.8
	5	1338.6	1553.7	1708.9	1425.1	1816.8	1452.4
	10	1431.7	1524.3	1589.6	1565.6	1764.0	1591.2
	20	1281.3	1490.7	1657.9	1377.7	1708.1	1805.0
30*30	0	955.4	1260.5	1263.1	1009.0	1326.4	1205.9
	5	1106.9	1317.5	1426.4	945.4	1204.4	1131.7
	10	988.3	1206.5	1222.1	1105.8	1125.2	1315.7
	20	966.5	1303.4	1374.0	1095.5	1336.7	1250.2
40*40	0	889.6	1219.7	1022.3	883.5	1098.7	1139.8
	5	923.7	1116.8	1205.0	891.9	1077.0	1098.3
	10	870.3	1101.1	1063.1	986.9	1112.4	1084.9
	20	908.8	1188.9	1245.4	894.9	1126.2	1045.9
50*50	0	918.8	1102.4	1179.4	933.0	1044.8	1077.6
	5	852.9	1039.7	1094.1	847.6	1026.2	1022.7
	10	883.2	1173.5	1177.3	884.5	1076.5	935.3
	20	808.6	1155.2	1219.5	881.0	1069.6	1153.9

4 Conclusion and Discussion

Based on the results above, we can conclude that performance of a matching task with the forefinger on a touch screen was mainly affected by button/icon size, and button iconography, digital or iconic. Button spacing did not significantly affect performance. When the button/icon was equal to, or bigger than 40*40 pixels, the reaction time was the fastest and the percentage of correct responses was the highest, around 98%. But the size 30*30 pixels also was a good choice for button/icon design if a reaction time of around 1200ms is acceptable for system efficiency. Button iconography did not affect the percentage of correct responses, but subjects responded to digits on buttons 200ms faster than icons or icon-digit combinations.

Almost all the subjects felt that the performance when wearing gloves was not as good as when they were not wearing gloves. The results contradicted this subjective impression. The data indicated that glove-wearing was not a problem for firefighters interacting with the computer touch screen except with the smallest buttons and spacing of zero.

The design of screen layouts for touch screen interfaces is always a trade off between the available screen space and the functionality that must be supported. This is particularly a problem for small screen displays. Our results indicate that the designer should attempt to maintain button size at a minimum of 30*30 pixels in favour of tradeoffs in spacing between buttons, which will have much less impact on performance than reducing button size.

References

1. Aftab, E.P, Sebastian, S., Tomescu, Milad, G.A.: Ishac: What visual information is used for navigation around obstacles in a cluttered environment? Canada Journal of Physiology and Pharmacology 82, 682–692 (2004)
2. Bay, S., Ziefle, M.: Children Using Cellular Phones: The Effects of Shortcomings in User Interface. Human Factors 47(1), 158–169 (2005)
3. Bender, Gregory, T.: Touch screen performance as a function of the duration of auditory feedback and target size, Ph.D.Dissertation, Wichita State University, 108 pages (1999)
4. Comerford, R.: Touch screen technologies for big size display. Electronic Products China 1, 4–10 (2005)
5. Yao, G.: Application of touch screen in HCI. Construction Machinery & Maintenance 10, 16–20 (2004)
6. Zhang, E., Zhang, A.: Development and application of touch screen technology. Journal of Shandong Normal University (Natural Science), 1 (2002)

Usability Evaluation of Children Edutainment Software

Danli Wang, Jie Li, and Guozhong Dai

Institute of Software, Chinese Academy of Sciences, Beijing 100080, China
danli@ios.cn

Abstract. Owing to its educational content and entertainment model, edutainment software is getting widespread interest. Given the fast increasing popularity of the computer, edutainment software shows a promising prospect of extensive development. However, just like the HCI software, children's edutainment software possesses the problem of usability. This paper, based on the study of usability evaluation methods, provides a scenario-based software evaluation method, and used the method to evaluate the Children Heaven edutainment software. Analysis of the evaluation results recognized the success of the software design. It also gives suggestions and comments for revising the software. The process of the evaluation experiments also verified the effectiveness of the scenario-based usability evaluation method for software.

Keywords: Software Usability, Scenario-Based Evaluation Method, Edutainment Software for Children.

1 Introduction

Because of the educational content and entertainment model, edutainment software is getting more and more attention. And it has a great development space under the computer's increasing popularity. There are many edutainment software products for children overseas [1-3]. But based on our review of the various kinds of edutainment software, there is no edutainment software fitting Chinese children and Chinese culture. So we develop an edutainment software—Children Heaven for Chinese children. Using pen-based multimodal interaction model and user interface for children, the software gives a garden for children gaming and studying with contents of combining education and entertainment.

Scenarios represent a good tool for communication [4]. Scenario-based design is a design methodology that considers scenarios as a central artifact in system design. It is an effective method to design an interactive system. Prof. Rosson proposed the scenario-based usability engineering model [5]. But there is not given detailed description which kind scenarios use and how use it in the software evaluation phase.

A scenario-based software evaluation method is proposed in this paper. Using this method, Children Haven is evaluated. An evaluation solution of usability for the edutainment software is designed and the evaluation experiment is accomplished. Some advice is given to improve system by analyzing experiment results. Moreover, the effectiveness of the proposed method is validated by experiment.

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 622–630, 2007.
© Springer-Verlag Berlin Heidelberg 2007

2 Related Works

2.1 State of the Art in Children's Edutainment Software

As computers become popular, more and more children begin to use computers. Children play games, chat with friends, tell stories, study history or math, and today this can all be done supported by new technologies. From the Internet to multimedia authoring tools, technology is changing the way children live and learn[6]. Research shows that the age of children using computers is decreasing. However, in the past such computer systems and software were all prepared for adults, and were not suitable for children. Therefore, researchers began in-depth study based on various aspects of children's applications. Thus many kinds of children's software were developed[1-3].

Edutainment software is educational entertainment software that is designed to entertain as well as help educate at the same time. This type of software may be a game that helps at the same time with a child's math skills [7]. Major universities and institutions that conduct research in children's edutainment software include University of Maryland's HCI Lab[6], MIT's Multimedia Lab[8],Carnegie Mellon University's HCI Research Institute[9], University of Tokyo's UI Research Group[10], University of Cambridge[11], and some other European Universities [12].

The key technologies of children's edutainment software involve knowledge in computer science, children's psychology, cognitive psychology, education, etc. Specifically, these include special-purpose design for computers [6,8], the inter-psychology between children and computers[6], children's understanding of human-computer interface[8], 3D models for children[10], programming for children[9], the impact of cooperation on children[13], among others. Besides, many theories, models and frameworks that have proven to be effective for adults also need further study and exploration to prove their suitability for children's applications.

Various research institutions working on children's hardware and software theories and systems' development and production have produced much useful work. But when compared to the research efforts in adult's software, there is still a lot of difference. Also much such work remains in the experimental stage. There are many domestic and overseas commercial children's application software packages. Essentially these applications attract children with beautiful images and multimedia entertainment methods. Yet these applications have not met children's needs, have not focused on children's psychological attributes as their guide for design and development. Essentially the traditional HCI means are employed.

China has nearly 400 million children, representing a very large market for children's software. As the popularity of computers and information technology grow, the computerization of education is unstoppable. This will cause more children to begin their exposure to information technology earlier. Therefore, the influence of edutainment software will become more and more important. However there exist these problems in China's education software: very little intelligent software, there almost does not exist education software that reflects the Chinese characters and Chinese culture, most children's software is Internet games. Such software is not suitable, both in contents and in formats, to the development of Chinese children. In general there is seldom self-owned copyright software.

Based on the afore-mentioned reasons, we developed the Children Heaven Edutainment software. This software adopted pen-based HCI technologies and virtual reality technologies. The interface design to the largest extent met the children's cognition and psychology. The content design is according to the intellectual development of children of various ages. Intellectuality and knowledge are the pronounced features. The software attempts to enhance the knowledge and skills of every child that visits Children Heaven.

2.2 Scenario-Based Design

Scenarios are some stories about people and their activities [4]. Let's see a concrete example: *A man wants to find a meeting agenda of June 20th, 2006. First he turns on his computer and invokes the window explore. Then he opens his routine file folder, in which he finds the meeting agenda folder. Finally, he locates and opens the agenda file in the agenda folder.*

Scenario-based design is a design methodology that considers scenarios as a central artifact in system design. The approach encourages user involvement in system design, provides shared knowledge and information among the people participating in the system development project, envisions the uncertain future tasks of the end users, and enhances ease of developing instructional materials [14].

The earliest successful application of scenario-based design was the voice message system developed by IBM for the L.A. Olympic Games [15]. Since then, the design found its way in various types of software development. In HCI, scenarios are used to describe detailed context to facilitate the design decision-making [4]. In software engineering, use cases are employed to depict the situations of a system in use, while scenarios are the instances of use cases [16]. Requirement engineering uses scenarios to record the observance and analysis by the users, from which the requirements are extracted. In addition, scenarios can also be used in requirement assessment [5]. The reference [17] introduced the concept of interface scenarios, and applied it into the PUI design.

3 The Children Heaven Edutainment Software

Children Heaven Edutainment Software is designed for children aged 3 to 12 years. The software's interface design fully considered the attributes of children to make it convenient and easy to use. It is very suitable for China's pre-school children and elementary school students' extracurricular self-guided learning and game playing.

The system uses virtual reality technology to construct gaming environments. It provides a newer operating model under virtual reality by using pen-based interface. The system focuses on developing the children's ability to observe, think, concentrate, imagine, and other non-intelligence factors. It uses cartoon-like 3D interface. Using dynamic and interesting stories, comics, games, music, and narratives, to guide the

children into mathematics, Chinese, nature, painting, handcrafting, music, and other learning, the system enhances the children's levels of cognition and knowledge. Children learn from playing, and play from learning. They develop interest in learning and potential for creativity, thus opening their unlimited ability to innovate.

Children Heaven consists of the function room, outdoor, and other scenery. Within the function room there are learning room, composing room and gaming room. Each room includes children's drawing board, learning words from pictures, puzzles, seven-piece board, simple notes composition shelf drums, and other software (a.k.a. playware); The outdoor includes the entrance, roads, grassland, rivers, trees, and other natural scenery. Figure 1 shows the system's functions.

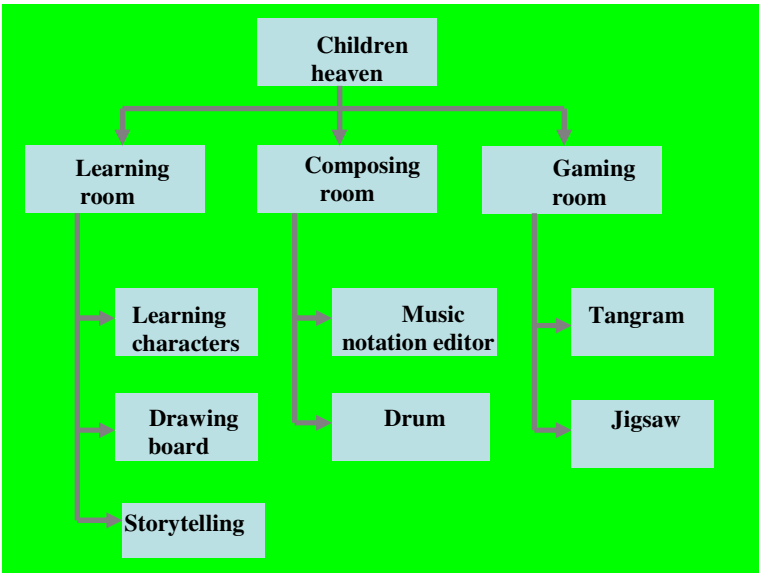


Fig. 1. Functions of Children Heaven

4 Software Scenario-Based Usability Evaluation Method for Software

4.1 Interface Scenarios

The interface scenario describes the process of user tasks in detail based on the interface in the form of pictures and texts. Figure2 shows a simple example. The following scenario describes creating a new meeting in the file management of meeting system. This kind of scenario does not only help the user understand the future system by giving him an intuitive description but also facilitate the communication between the designers and the users.

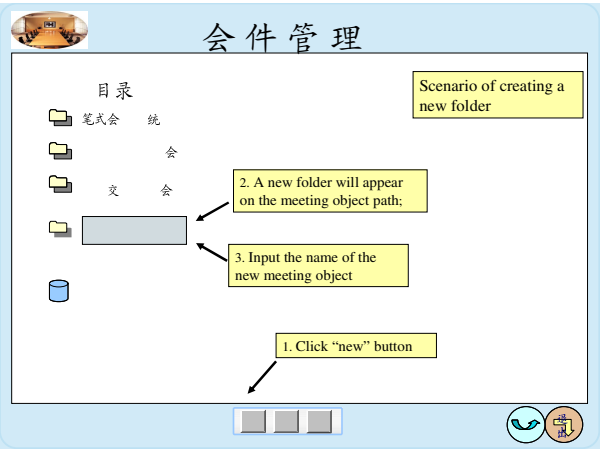


Fig. 2. Interface scenario

4.2 Scenario-Based Usability Evaluation Method for Software

Scenario-based software usability evaluation method is a method that uses multiple scenarios to evaluate software. The method includes three stages: training, tasting and analysis stage. In the training stage, the interface scenarios is used to design the training document and test missions, and let user understand the software usage, and learn it in the short time. In the real testing, video scenarios and observation scenarios is used to record and analyze operation conditions, to find the problems of the software, and the preference of the users. During the analysis stage, according to the testing data and investigation result, scenarios for analysis charts and tables are made, to describe the test results, analyze the scenarios, and to combine video scenarios and surveys, thus providing suggestions for software revisions.

5 Usability Evaluation for Children Heaven Edutainment Software

5.1 Evaluation Method

Based on massive volume of user study and mission analysis, we designed the evaluation solution for Children Heaven edutainment software using scenario-based usability evaluation method. We focused on the roaming of 3D environments, functions and intelligent contents of playware in evaluating the systems. We first designed the evaluation method, mainly including the software introduction, training documents, test tasks and survey papers, etc. The training documents and test tasks are both designed with interface scenarios. Within the training documents we designed children's interface scenarios for the roaming and those used by various functional software. For the testing tasks, we also provided specific task s for scenario descriptions. Due to page limit, we present Figures 3 and 4.

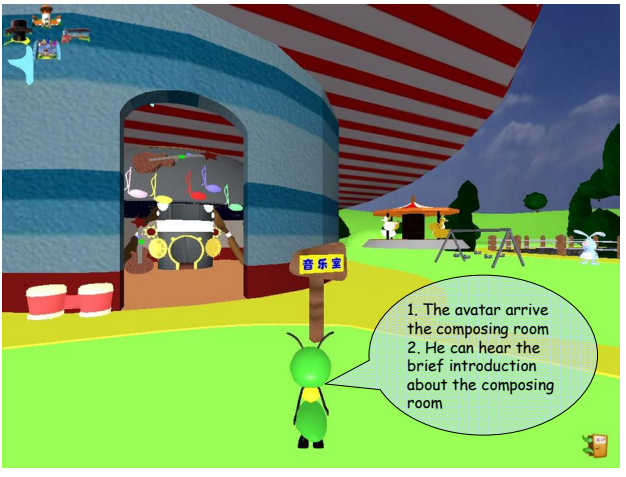


Fig. 3. The roaming scenario

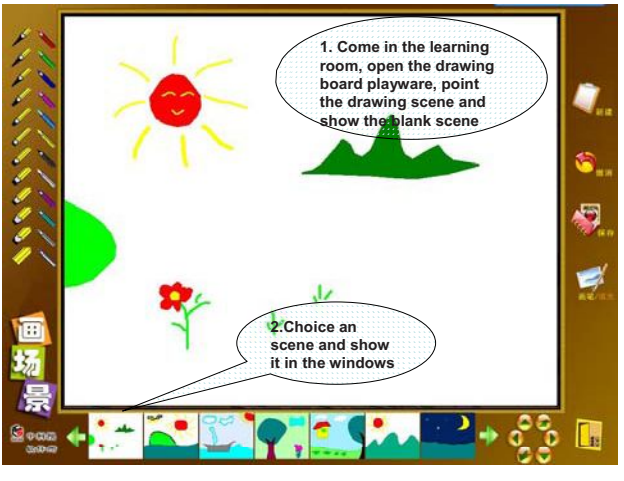


Fig. 4. Children's drawing board painting selection scenario

5.2 Evaluation Experiments and Analysis

11 persons took part in testing the software usability. Five testing tasks are set: (1) roaming to the music room under 3D environment; (2) enter learning room's children drawing board playware, and paint the required patterns; (3) using the learning word from pictures playware, imitate a Chinese character; (4) enter the entertainment room, open the puzzle playware, and start partitioning and reassembling the puzzle; (5) after the subjects finished training, let the subjects become familiar with the software, then start the tests. Test figures and test scenarios are recorded.

After finishing test, all testers answered the same questionnaires. The dimensions of questionnaire include 3 aspects: functions, contents and 3D interface design. The synthesis of the above items reflects the effectiveness, efficiency and user satisfaction. According to analyzing the questionnaires, the analysis results are listed in the following table1-3. From the table, the user gives a good evaluation on the easy learnability, naturalism, the usability of the interface interaction, and level of interest. Meanwhile, other indexes are more than the average score (the average score is 3).

Table 1. Average score and standard deviation on each evaluation dimension of function usability (min: 1, max: 5)

Function evaluation	Average Score	Standard deviation
Usability	3.61	0.69
Learnability	4.21	0.81
Reliability	3.85	0.66
Natural	4.18	0.89
Interface interaction	4.30	0.78
Overall evaluation	3.55	1.04

Table 2. Average score and standard deviation on each evaluation dimension of content (min: 1, max: 5)

Content evaluation	Average Score	Standard deviation
Knowledgeability	3.73	0.90
Reasonableness	3.36	0.85
Entertainability	4.00	0.82

Table 3. Average score and standard deviation of interaction design usability (min: 1, max: 5)

HCI design	Average Score	Standard deviation
Ease of use	3.39	0.98
3D UI design	3.45	0.93

Although the overall evaluation of the system is satisfactory, the system has some problems still. Several suggestions of improving the system are proposed combined with the survey of the evaluation process and the analysis of the testing results. For example, under 3D environments the user always has problems finding the designated function room, especially when using the short-cut methods. The user cannot discern where the destination is. To improve this, we provide a navigational map, and mark every map artifact with its name for the short-cut method.

In general, the evaluation test regards the overall condition of the current systems as positive, and it also finds the existing problems, which will provide the detail aim for improving the system. In addition, it verified the effectiveness of scenario-based evaluation method. Furthermore, from the test results and analysis, we claim that scenario-based software usability evaluation method is an effective method.

6 Conclusion

Usability is a prevalent problem for software. Children's intelligent software is no exception. This paper, using scenario-based usability evaluation method, evaluated the usability of Children Heaven edutainment software to improve the usability of the software so as to make it more suitable for Chinese children. Through our evaluation experiments, we discovered the software's many problems, and provided suggestions for revisions. However, usability evaluation requires more than one round of evaluation to solve all the problems. It is best to repeat this process several times to provide more usable software.

Acknowledgments. The authors appreciate the evaluation performed by Tiegang Cheng, Jie Zhang, Dehua Cui, Guangyu Wu, and the implementation of the Children Heaven by Fengjun Zhang, Lianen Ji. The research is supported by National Grand Fundamental Research 973 Program (Grant No.2002CB312103) and the National Natural Science Foundation of China (Grant No.60373056).

References

1. Bers, M., Cassell, J.: Interactive Storytelling Systems for Children: Using Technology to Explore Language and Identity. Journal of Interactive Learning Research 9, 603–609 (1999)
2. Hourcade, J.P., Bederson, B.B., Druin, A., Taxen, G.: KidPad: Collaborative Storytelling for Children, http://www.cs.umd.edu/hcil/kiddesign/hourcade-kidpad-paper.pdf
3. Alborzi, H., Druin, A., Montemayor, J., Sherman, L., Taxén, G., Best, J., Hammer, J., Kruskal, A., Lal, A., Plaisant Schwenn, T., Sumida, L., Wagner, R., Hendler, J.: Designing StoryRooms: Interactie Storetelling Space for Children. In: Proceedings of Designing Interactive Systems. In: Proc. ACM Desiging Interactive Systems (DIS'2000), NY, pp. 95–105 (2000)
4. Carroll, J.M.: Scenarios and Design Cognition. In: Dai, G.Z. (ed.) Dong, S.H, Chen, Y. D., Ren, X.S. (sub-ed.) Proceedings of the APCHI, User Interaction Technology in the 21st Century, China, Beijing, Science Press, vol.1, pp. 23-46 (2002)
5. Rosson, M.B., Carroll, J.M.: Scenario-Based Design. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook, pp. 1032–1050. Lawrence Erlbaum Associates, New Jersey (2002)
6. Druin, A.: The Role of Children in the Design of New Technology. Behaviour and Information Technology 21, 1–25 (2002)
7. http://www.computerhope.com/jargon/e/edutainm.htm
8. Bers, M., Cassell, J.: Storytelling Systems: Constructing the Innerface of the Interface. In: Cognitive Technologies Proceedings '97, pp. 98–108 (1997)
9. Pane, J.F.: Designing a Programming System for Children with a Focus on Usability. In: Proceedings of CHI 98, Los Angeles, pp. 62–63 (1998)
10. Igarashi, T., Matsuoka, S., Tanaka, H.: Teddy: A sketching interface for 3d free from design. In: Proceedings of SIGGRAPH 99, pp. 409–416 (1999)
11. Rode, J.A., Stringer, M., Toye, E., Simpson, A.R., Blackwell, A.: Curriculum focused design. In: Proceedings ACM Interaction Design and Children, pp. 119–126 (2003)

12. Höysniemi, J., Hämäläinen, P., Turkki, L., Rouvi, T.: Children's Intuitive Gestures in Vision Based Action Games. Communications of the ACM 48, 44–50 (2005)
13. Druin, L.: Cooperative Inquiry: Developing New Technologies for Children with Children. In: Proceedings of CHI'99, pp. 592–599 (1999)
14. Go, K., Carroll, J.M., Imamiya, A.: The Blind Men and the Elephant: Views of Scenario-Based System Design. Interactions 6, 45–53 (2004)
15. Jacobson, I., Christersson, M., Jonsson, P., Overgaard, G.: Object-Oriented Software Engineering: A Use-Case Driven Approach. Addison-Wesley, Reading, MA (1992)
16. Potts, C., Takahashi, K., Anton, A.I.: Inquiry-based Requirements Analysis. IEEE Software 11, 21–32, 3 (1994)
17. Wang, D.L., Hua, Q.Y., Dai, G.Z.: Research on User-Centered Scenario-Based Design. Chinese Journal of Computers 28, 1043–1047 (2005)

Effect of Different Modal Feedback on Attention Recovery

M.C. Whang[1], H.J. Hyun[2], J.S. Lim[1], K.R. Park[1], Y.J. Cho[1], and J.S. Park[3]

[1] Division of Media Technology, Sangmyung University, Seoul, Korea
7 Hongji-dong, Jongro-gu, Seoul, Korea
{whang,jslim,parkgr,ycho}@smu.ac.kr
[2] Brainware Research Center, Sangmyung University, Seoul, Korea
7 Hongji-dong, Jongro-gu, Seoul, Korea
systers@paran.com
[3] Electronics and Telecommunications Research Institute, Taejon, Korea
161 Gajeong-dong, Yusong-gu, Taejon, Korea
parkjs@etri.re.kr

Abstract. This study aims to empirically examine the effect of feedback on attention recovery. The role of feedback has been proven to be positive in particular to extend the limitation of attention resource. We studied the impact of feedback on attention by varying its type and modality. An experimental system was developed to observe how accurately the participants performed the figure-matching task with differential feedback provided on a real-time basis based on the ADHD (attention deficit hyperactivity disorder) diagnostic model. Investigated in this study were two types of feedbacks (1) single feedback such as visual, auditory and tactile stimulus and (2) double feedback with two types of feedback. Eight university students participated in this study with six different feedback conditions and controlled conditions. The results showed that the tactile feedback and the combined tactile with visual feedback significantly contributed most to attention recovery and performance.

Keywords: attention, feedback, modality, attention recovery.

1 Introduction

Attention is important human factor having close relationship with performance together with motivation, and cognitive ability in learning and information processing [1]. Attention has been defined as the process of psychological concentration to the incidence which people can perceive and feel [2]. Also, attention includes the meaning of alertness, arousal, sustained attention, selective attention and resource as well as processing capability [3]. Human performance has been tried to keep good quality using optimal range of attention. Inverse U pattern between attention and human performance has been determined by Law of Yerkes-Dodson. In his model, the relationship between attention and performance was explained. Not only attention improves performance but also the optimal range of attention provides maximum performance. He has shown clue of improving human performance by controlling

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 631–636, 2007.
© Springer-Verlag Berlin Heidelberg 2007

attention. However, the optimal attention needed to be defined according to type of task and individual difference of human characteristics.

There have been various attempts for the improvement of attention with a purpose of enhancing people's task performance capability while reducing mistakes. Recently, a model explaining three stages where the level of attention reaches to performance more delicately was presented [4]. According to this model, arousal occurs at the emotional level (affective arousal) in the first place. At this time amygdale of brain is mainly involved and heart activity becomes faster physiologically, which we can define as a sort of physical excitement state. In the second stage, arousal is converted to effort. This is the process where hippocampus is mainly involved and here emotional arousal of the previous stage is converted into cognitive level. Accordingly, state of brain is changed into active state in which alpha wave in the brain wave is decreased. Last stage is the state of preparation for activity. This is the stage where physical preparation for the participation in the particular activity to which amygdale is mainly involved is completed. Therefore, attention can be perceptual (affective) and cognitive having different physiological response. However, less clear is effect according to different attention on human performance.

Feedback according to low attention state has generally been accepted as one of the methods for recovering or maintaining the resource of attention [5]. NASA has been developing bio feedback mechanism called "SMART Brain Game System" based on EEG which can determinate the level of arousal. They used the feedback of varying difficulty of game control at low attention. Still, quality of feedback has not clearly stated in terms of feedback effect. Therefore, this study assumed attention is improved by feedback and did differently according to feedback of different modality. This study was attempted to analyze empirically the impact of feedback on attention in terms of human performance by varying feedback type and modality such as visual, auditory, tactile single feedback and double feedback combined them.

2 Methods

Eight subjects (four female and four male) participated in this study and did not have any perceptual problem. Their ages were averaged to twenty nine. For the purpose of measuring the level of attention, a computerized measurement system has been developed using the model of ADHD (attention deficit hyperactivity disorder) clinical diagnosis [7]. This model required the matching task of identifying figures together as shown in Figure 1. One of four irregular figures presented on the screen for 0.1 second to be identified from four different types of figures presented continuously. The participants should click the mouse button when they were matched. Attention level was considered high at right click but low at wrong click or at missed click. The participants experienced feedback at low attention.

Three modalities consisted of visual, auditory and tactile stimulus, were used for feedback. Visual feedback was cognitive showing the correct number during matching task. Auditory feedback was perceptual using sound and tactile feedback was also perceptual generating the vibration at mismatching. Using three modalities were made up seven conditions composed of controlled feedback, three single modal feedbacks and three double modal feedbacks making pair of three conditions.

The participants experienced the feedbacks before experiment and set contrast and brightness of visual stimulus, volume of auditory stimulus and frequency of tactile stimulus from their subjective preference. They performed figure-matching task with each feedback condition for 200 seconds. A participant was asked two hundred tasks in seven feedback conditions. Two hundred tasks were repeated seventy times. Therefore, the collected data was total fourteen thousands. The difference of feedback effect according to different modality was to find within a subject. The sequence of feedback was randomized for avoiding order effect. The number and reaction time of respective matching, mismatching and missing matching were measured.

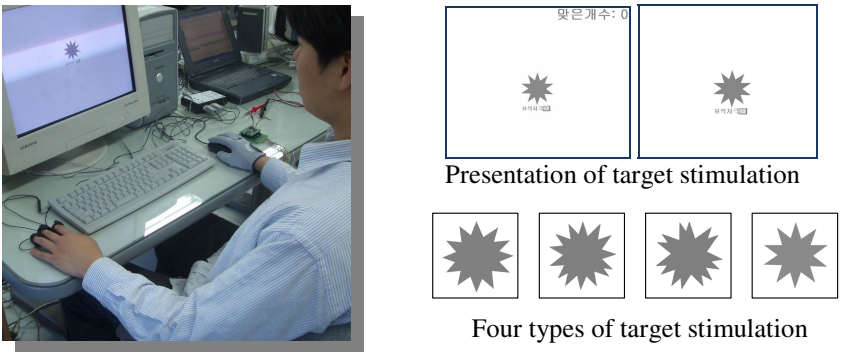


Fig. 1. Shown is matching performance with feedback including visual, auditory and tactile sensation

3 Analysis

Correlation between test frequency and reactions including correct, failed and missing was analyzed to find training effect. Two hundred tests were averaged and eighty repentances were listed for independence test among reactions. The results as shown in Table 1 indicated low correlated with test frequency. Therefore, no training effect was proven. Means of feedbacks using Kruskal-Wallis H was compared with performance. Rank of performance was analyzed according to different modal feedbacks which were visual, auditory, tactile, visual with auditory, visual with tactile, and auditory with tactile.

Table 1. Correlation between test frequency and reactions

	Correct Reaction	Failed Reaction	Missing Reaction
Pearson Correlation	.092*	-0.114**	-.070

*. Correlation is significant at the 0.05 level

**. Correlation is significant at the 0.01 level

4 Results

This study assumed the number of matching were the index of attention based on ADHD Diagnostic system [6]. The differences of attention improvement in the condition of seven feedback conditions were analyzed based on mean rank of the reaction time and the performance defined as equation (1). There were significant difference in the performance and significant difference in the reaction time among seven feedback conditions as shown Table 2. Also, significance is shown in correct reaction, failed reaction and missing reaction according to seven feedback conditions.

$$\text{Performance} = \text{Correct Reaction} - \text{Failed Reaction} - \text{Missing Reaction} \quad (1)$$

Compared with the controlled condition meaning no feedback, all the feedback increased the correct reaction and decreased the missing reaction. However, the failed reaction was decreased only with the tactile feedback. Other feedback did not help to decrease error of performance. Reaction time got fast with all kinds of the feedback. Therefore, the feedback showed the effect on improving human performance and productivity.

There was different feedback effect on performance and reactions among modalities. The tactile feedback of single modality drove best performance, most correct reaction, least failed reaction and least missing reaction among three single feedbacks. Compared with the visual feedback in single modality, the auditory feedback showed better effect. Double modal feedback of the visual and the tactile feedback led strongest effect on correct reaction, failed reaction, missing reaction and performance among three double feedbacks. Double modal feedback of the visual and the auditory showed least effect on the reactions and the performance.

Table 2. Mean rank of task under the seven feedback conditions

Dependent variables / Independent variables	Feedback Conditions							Asymp. Sig.
	Controlled	Single Modality			Double Modality			
		A	B	C	A x B	B x C	C x A	
Correct Reaction	236.58	261.92	277.86	303.71	271.19	279.46	339.52	0.038
Failed Reaction	262.97	268.49	278.43	254.67	294.44	271.61	318.77	0.031
Missing Reaction	320.44	298.77	279.45	249.20	282.85	273.08	221.06	0.026
Performance	241.75	268.21	279.10	309.97	275.27	285.42	307.28	0.048
Reaction time	318.90	292.21	278.45	263.13	289.04	265.68	248.75	0.046

* A: Visual Feedback, B: Auditory Feedback, C: Tactile Feedback

Similar pattern was shown in the reaction time. The reaction time at matching was fastest in the tactile feedback condition and in the double modal feedback combined the visual and the tactile but slowest in the visual feedback condition and in the double modal feedback combined the visual and the auditory. Therefore, showing most improved the reactions and the performances was the tactile feedback in the single modality and the combined visual and tactile feedback in the double modality. The comparison of single modal feedback with the double modal feedback showed the double modal feedback was more effective as shown Table 3. Considering comparison with the controlled reaction showing improved failed reaction only with tactile feedback and improved other reactions, the combined feedback effect was attributed to main contribution from the tactile feedback.

Table 3. Mean rank of task under the single and double modality feedback

Dependent variables / Independent variables	Feedback Conditions			Asymp. Sig.
	Controlled	Single Modality	Double Modality	
Performance	241.75	284.42	288.55	0.018
Reaction Time	318.90	278.77	268.83	0.017

5 Conclusion and Discussion

This study analyzed the feedback effects on attention according to different modalities. Computerizing ADHD test model was developed to measure matching task automatically. A subject were asked to experience seven feedback conditions consisted of three single modal feedback which were visual, auditory and tactile and three double modal feedback combined three single feedbacks. Eight subjects had two hundred matching tasks with eighty repentance and their measurements were the number and the reaction time of matching. This study found the feedback of any modality improved attention and performance. Also, there was different effect with the feedback of different modality. The results compared the single feedback with the double feedback showed that double feedback drove better performance and faster reaction. Also, modality combining was specified to visual and tactile stimulus. A participant should recognize the number of matching during visual feedback but just feel during auditory and tactile feedback. Therefore, the visual feedback was cognitive and the others were perceptual. The results indicated that the combined feedback of cognitive with perceptual stimulus led better performance than the combined of perceptual feedbacks. Combining pattern of feedback using different modalities may result in triggering better human performance. In the case of using single modal feedback, the tactile feedback drove to improve human performance most effectively. Interestingly, the visual feedback related the performance was less helpful for enhancing or maintaining attention than other perceptual feedbacks unrelated to performance. According to Back and Boucsein's model [4], first stage of

attention is emotional and resulted from activation of autonomic nervous system. This stage is perceptual rather than cognitive. The task in this study was instantaneous so that attention was induced in first stage of Back and Boucsein's model. Therefore perceptual feedback should be more effective than cognitive in short period of task. Orienting reflex is the attention concentration response for a strange stimulus instead of voluntary concentration of attention [7]. In the state of orienting reflex, limbic system causes the excitements of sympathetic nerve system as well as parasympathetic nerve system simultaneously. This orienting reflex state is very temporary and unstable state physiologically. The balance in such state is broken and the state is converted into the situation where the state will be led by either one of the sympathetic nerve system or parasympathetic nerve system in short period of time [8]. This explains why attention level can be emotional or perceptual for a short time. Considering the results of this study may give some human factors for improving human performance in working conditions controlling attention. Perceptual feedback keeps better attention and performance.

Acknowledgement. This study was supported by the project titled "Five Senses Information Processing Technology Development for Network Based Reality Service" funded by Ministry of Information and Communication Republic of Korea.

References

1. Rizzo, A.A., Buckwalter, J.G., Neumann, U.: Virtual reality and cognitive rehabilitation: a brief review of the future. Journal of Head. Trauma Rehabilitation 12(6), 1–15 (1997)
2. Solso, R.L.: Cognitive psychology, 4th edn. Allyn and Bacon, London (1995)
3. LaBerge, D.L.: Attention. Psychological Science 1, 156–162 (1990)
4. Back, R.W., Boucsein, W.: Engineering Psychophysiology: Issues and Applications. Lawrence Erlbaum, Mahwah, NJ (2000)
5. NASA: S.M.A.R.T. brain game system (2005) http://www.sti.nasa.gov/tto/spinoff 2003/hm_2.html
6. Shin, M.S., Jo, S.J., Jeon, S.Y., Hong, G.E.: Research on the Development and Standardization of Computerized Diagnosis System for the impairment in Attention. Infant and youth mental medical science 11(1), 91–99 (2000)
7. Sokolov, E.N.: The Orientating Reflex: The "Targeting Reaction and Searchlight of Attention. Neuroscience and Physiology 32(4), 347–362 (2002)
8. Kimmel, H.D.: Habituation and dishabituation of the human orienting reflex under instruction - induced stress. Physiological psychology 13(2), 92–94 (1985)

Do We Talk Differently: Cross Culture Study on Conference Call

Xingrong Xiao, Chen Zhao, and Shaoke Zhang

IBM China Research Lab, Bldg 19, Zhongguancun Software Park,
Beijing, P.R. China, 100094
{xiaoxr,zhaochen}@cn.ibm.com, shaoke.zhang@gmail.com

Abstract. Cross cultural collaboration is popular in the world with increasing globalization, where cultural issues are important to be explored. In this paper, we reported an investigation of culture differences and cultural effects on communication problems in cross culture conference call using an ethnographic technique which refers to long interviews. In these interviews, communication differences among Chinese, Japanese, Indian and Americans were investigated. Our results showed that (1) culture differences in conference call existed in the dimensions of indirectness, power distance, assertiveness, language and speaker-centered vs. listener-centered; (2) and these culture differences caused communication problems in conference call such as misunderstanding, bad impression, unequal participation.

Keywords: Culture, Communication Style, Conference Call.

1 Introduction

In the world with increasing globalization, more and more corporations undertake global projects. Cross cultural collaborations become indispensable and ubiquitous in business and workplace. These changes bring new challenges and require considerable attention because of culture crashes, and there is increasing concern about how cultural differences may affect work-team performance.

Researchers investigated the effects of culture and diversity on performance of distributed group collaborations. Results of these studies showed that the relationship between performance and a team's cultural composition had been mix. While some studies found that multicultural teams often suffered from greater group process losses such as higher conflict, less participation, than culturally homogeneous team [1], others revealed that multicultural teams could perform as well as homogenous teams [2].

From above researches, it can be seen that study results are not the same and further studies are needed. In this study, we investigated culture differences in communication styles in cross cultural meetings supported by teleconference which is always called conference call. Meanwhile, it was examined if these culture differences would cause problems or not as well as what kinds of problems would be produced by a certain difference. We selected conference call, which was a telephone call that interconnected three or more phones simultaneously, as the typical scenario

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 637–645, 2007.
© Springer-Verlag Berlin Heidelberg 2007

for two reasons. First, it was a common and ubiquitous channel for cross cultural collaboration. Second, it excluded the factors like body language—while we admitted that body language was also important in cross cultural communication, we tried to focus on the communication styles reflected in the meeting contents.

2 Culture Study on Communication Style

Norton regarded communication style as "the way one verbally, nonverbally, and Para verbally interacts to signal how literal meaning should be taken, interpreted, filtered, or understood" [3]. As communication is important in daily life, communication style has been a hot topic which attracts researchers' attention [4, 5, 6, 7, 8, 9, 10, 11]. They investigated the way people from different cultures expresses and talked. Indirectness, Assertiveness, context, and speaker/listener-centered are four aspects of communication styles which researchers have paid most attention to.

Indirectness

Indirectness has been shown to vary between cultures. Indirect expressions were more common in collectivist culture because the emphasis on face-work [5] and people from collectivist cultures (such as Koreans) were more indirect than people from individualistic cultures [6, 7].

In other similar aspect, Hall reported that there was more indirectness in high-context cultures than in low-context cultures [8]. In high-context cultures communications are more subjective and multilayered, colored by relationships, history and status; in contrast, in low-context cultures, events have single and universally understood objective meanings [4].

Assertiveness vs. Responsiveness

Richmond and McCroskey developed the Assertiveness-Responsiveness Measure. Assertiveness reflects a person's willingness to speak up for her- or himself in interaction or let others take advantage of her / him; responsiveness involves being other-oriented, considering others' feelings, and listening to what others say [9].

Klopf, Ishii, & Cambra conducted a survey finding that the Japanese was in a moderate level of assertiveness while the Americans showed a high degree of assertiveness, which meant it was relatively easy for an American to make a request and actively disagree with another's opinion and express his or her personal rights and feelings, whereas the Japanese found it was not so easy to do the same [10].

High-context communication vs. low-context communication

High-context and Low-context Communication refers to the degree to which speakers rely on factors other than explicit speech to convey their messages [12]. In high-context communication, communicators share much information and context in which they are talking about, while in low-context communications, communicators understand each other relying on words they say.

Gao and Toomy pointed out that Chinese communication was a high-context communication, as Chinese always emphasized and identified with some expressions such as "yan bu jin yi" (not saying all that is felt) and "yan wai zhi yi" (more is meant than meets the ear) [11].

Teruyuki Kume etc. [10] compared the communication styles among Japanese, Americans and Chinese by 5 TV dramas aired in the 1990s in Japan, the U.S. and

China. Results revealed that American communication was low-context while Japanese and Chinese communication was high-context.

Speaker-centered and listener-centered

Teruyuki Kume etc. revealed that Americans were speaker-centered while Chinese and Japanese were more listener-centered. For example, Americans often proceeded with the discussion by requesting information, making confirmation, which were rarely found in Chinese and Japanese drama. Americans were eager to talk, expressing their opinions freely while Chinese and Japanese listeners rarely asked questions but just followed with short expression of support [10].

3 Method

3.1 Participants

Six participants were investigated in the study. They were Chinese who do research and management work in a large multinational corporation. Two of them have lived in American for more than 20 years. The rest were native Chinese, who had rich experiences in cross cultural communication with foreigners like the American, Japanese, and Indian, averagely 12 years with the standard deviation of 8. They participated in conference calls at least once a week, at most 4 or 5 times a day, and each conference call lasted about 1 hour. Their meeting activities included discussing technique problems, reporting and tracking status, negotiating and decision making. In all of the meetings, they used English as the main language.

3.2 Procedure

The study was based on semi-structured interviews. Each participant was interviewed separately face-to-face for about one hour. In the interview, we investigated culture differences in communication styles and their influences on outcome and efficiency in conference calls. All data were recorded by tape for following transcription.

3.3 Data Coding

As previous research revealed that culture differences in communication styles existing in the aspects of indirectness, assertiveness, high-context vs. low-context communication and speaker-centered vs. listener-centered communication. The transcriptions were coding according to these four aspects.

All transcriptions were coded by 2 researchers, and the coherence is 92%. The disagreement was discussed to be determined.

4 Results

4.1 Culture Differences in Conference Call

Participants pointed out that culture differences in conference call existed in the dimensions of indirectness (6/6[1]), power distance (6/6), assertiveness (5/6), language

[1] 6/6 means 6 of the experts considered there were culture differences in the dimension of Indirectness; the rest may be deduced by analogy.

(4/6), speaker-centered vs. listener-centered communication (4/6) and high-context communication vs. low-context communication (2/6).

In addition, all of the participants (6/6) reported that power distance had great effects on the way people talk.

In the study, participants mainly reported culture differences in communication styles among Americans, Chinese, Indian and Japanese. In the following, culture differences in concrete communication styles were described.

Indirectness. Result revealed that Americans were direct, Chinese and Japanese were indirect and Indians were in the middle. Table 1 showed the concrete characteristic of how Americans, Indians, Chinese and Japanese talked indirectly in conference call as well as ratio of participants mentioned the characteristic. It can be seen all of the participants pointed out American was direct and Chinese was indirect.

Table 1. Characteristic of communication styles in Indirectness

Nation	Communication Characteristic	Ratio of participants mentioned the Characteristic
American	Express their opinions directly	6/6
	Be straightforward	1/6
	Say "NO" directly	1/6
	Criticize and praise directly	1/6
Indian	Be direct in work but their real feelings is difficult to be understood	1/6
Chinese	Express their opinions implicitly	6/6
	Express disagreement indirectly	4/6
	Threw out many reasons but hoped the other side to give an expected conclusion	1/6
	Seldom say NO directly	1/6
	Talk in a roundabout way	1/6
Japanese	Never say bad though dislike	1/6
	Always say good	1/6

Assertiveness. Table 2 showed the characteristics of communication style on assertiveness mentioned by participants. From the table, we could see that American was most assertive.

Speaker-centered vs. Listener-centered. Results revealed that American and Indians were speaker-centered while Chinese and Japanese were listener-centered in cross cultural meeting. Table 3 showed the detailed description.

Table 2. Characteristic of communication style on Assertiveness

Nation	Communication Characteristic	Ratio of participants mentioned the Characteristic
American	Speak up their ideas and questions immediately	4/6
	Insist on their own opinions	2/6
Indian	Tend to lead a discussion	1/6
	Be aggressive	1/6
	Be tough in argument	1/6
Chinese	Dislike to express their requirements	3/6
	Dislike to ask questions in meeting	2/6
	Don't know how to say "NO"	2/6
	Avoid conflicts	1/6
Japanese	Be quiet and speak up in key points	1/6

Table 3. Characteristic of communication style in Speaker-centered vs. Listener-centered

Nation	Communication Characteristic	Ratio of participants mentioned the Characteristic
American	Talk much	3/6
	Use a lot of "I" in speech	2/6
Indian	Spend a lot of time to speak	1/6
	Talk ceaselessly	1/6
	Take a lot of time to speak	1/6
Chinese	Use more "we" in their speech	3/6
	Dislike talking and tend to listening	3/6
	Be quiet in meeting	1/6
Japanese	Be quiet in meeting	1/6
	Talk less	1/6
	Tend to listening	1/6

High-context vs. low context communication. Only 2 of the 6 participants pointed the culture differences in High-context and Low-context communication. Table 4 showed the details.

Effects of Power distance on communication. Power distance didn't belong to communication style, but results showed that it had an important effect on communication. Table 5 showed the detailed communication characteristics caused by power distance.

Table 4. Characteristic of communication style in High-context vs. Low context communication

Nation	Communication Characteristic	Ratio of participants mentioned the Characteristic
American	Use a lot of explanatory words	1/6
	What they speak up is what they want to say	1/6
Chinese	Use simple words such as "Yes" or "No" to answer questions	1/6
	Have implication in words	1/6

Table 5. Characteristic of communication caused by power distance

Nation	Communication Characteristic	Ration of participants mentioned the Characteristic
American	Participate in discussion equally	3/6
	Common to express different opinions from managers	1/6
Chinese	Regard managers' voice as more important	2/6
	Pay more attention to authority	2/6
	Managers saying and employees listening	1/6
	Assume that managers are always right	1/6
	Let decision made by management team	1/6
Japanese	Pay more attention to authority	2/6
	Managers' voice can only be heard in meetings	1/6
	Try to fulfill managers' will	1/6

4.2 When Oriental Meets Occidental: Problems in Interaction

The data demonstrated that culture difference in such communication styles as indirectness, power distance, assertiveness and speaker-centered vs. listener-centered communication caused several communication problems which affected meeting satisfaction and efficiency.

Problems Caused by Indirectness. Indirectness was an important culture difference in conference call which was described above. American and Indian were less indirect

than Chinese and Japanese. Study data suggested that when these two kinds of people had meetings together, the communication problems were as follows:

(1) Misunderstanding
4/6 participants reported that culture differences in indirectness caused more misunderstandings in cross culture conference call than that in intra-national meetings. For example, Americans always gave a direct challenge only referred to business affairs. But Chinese maybe regarded it as negative evaluation and felt uncomfortable. In addition, Chinese didn't like saying "no" because of the importance of face, and it also made Americans misunderstood their real intention.

(2) External meeting
1/6 participants suggested that external meeting was more common in cross cultural meeting than in intra-national meeting because Chinese's indirectness often created some unrelated threads. For example, when in discussion, Chinese would like to give reasons first and hoped the other side to give a decision he expected. However, if the other side was direct, they always could not aware Chinese purpose, and should ask some questions about the reasons, so meeting subject would be shifted to the reasons, which in fact is not to be solved in the meeting. (This example was given by a participant.)

Problems Caused by Power Distance. Culture differences in Power Distance caused problems more often such as "some opinions can not be fully put forward" (5/6) and "unequal participation" (4/6). For example, if Japanese participated in a meeting, you could only hear managers voice; we didn't know what other Japanese thought about. However, managers didn't always know detail information, so opinions could not be fully put forward. (This example was given by 2 participants.)

Problems Caused by Language Use. All participants (6/6) pointed out that language was a big problem in conference call. First, in most conference calls, English is a working language, but for some of the attendees, English is not their native language which will influence on their expressions. Second, misunderstanding was sometimes caused because some attendees didn't know well about the slang, idioms, convention and so on.

Other Problems. Culture differences in dimensions of Assertiveness and Speaker / Listener Centered also caused communication problems such as 'bad impression' (3/6), 'pended questions' (4/6) and "unequal participation" (3/6). Bad impression meant if a participant in a cross cultural conference call who was listener-centered or less assertive would be regarded as uninvolved, inactive or less contributions. In addition, people who were not such assertiveness disliked asking questions; if they had any questions, they would like to think by themselves after meeting. Because of that they often lost the opportunity to clarify questions which may cause pended questions and finally affected meeting efficiency.

4.3 Measures to Avoid Culture Shock: Culture Awareness

In the end of the interview, each participant was asked how to avoid problems caused by culture differences in conference call. All of them said culture awareness was very important to reduce culture shock. They pointed out that in one hand if participants in conference call didn't aware the culture difference, they could not do anything to intervene it. And in the other hand, communication problems caused by culture differences would be decreased with the increase of cross cultural communication as people knew more about other culture.

5 Conclusions and Future Research

In this study, culture differences in communication style among American, Chinese, Indian and Japanese were investigated as well as communication problems caused by these differences in cross cultural conference call. Results showed that in cross cultural conference calls, people talked differently, especially in the aspects of indirectness, assertiveness, power distance, high-context vs. low-context communication and speaker-centered vs. listener-centered communication, which were consistent with previous study [7, 8, 9, 10, 11, 12,13]. In addition, participants pointed out that there were several culture differences in language use such as different use of slang, idioms, acronyms, and conventions. In this study, we also found that different communication problems were caused by specific communication styles, for example, indirectness always caused "misunderstanding" and assertiveness tended to cause "bad impression". According to these results it was found that to find the relationships between culture differences and communication problems was important to facilitate cross cultural communication.

Though culture differences and communication problems were found through the present study, however, in this study, we only used a single interview technique, which probably could not distinguish participants' perceptions on what they saw and listened from those on what they really experienced in very well. So in the future study, we would like to conduct a field study to observe culture differences and communication problems in real cross cultural conference calls to validate if culture differences existed in aspects of indirectness, power distance, assertiveness and speaker-centered vs. listener-centered communication. And meanwhile, another purpose of future study was to investigate problems caused by such culture differences as well as intervention to avoid culture shock.

References

1. Adler, N.J.: International Dimensions of Organizational Behavior, 2nd edn. PWS-Kent Publishing Company, Boston (1991)
2. Anderson, W.N., Hiltz, S.R.: Culturally Heterogeneous vs. Culturally Homogeneous Groups in Distributed Group Support Systems: Effects on Group Process and Consensus. In: Proceedings of the 34th Hawaii International Conference on System Sciences, pp. 1–14 (2001)

3. Norton, R.: Communicator Style: Theory, applications and measures. Sage, Beverley Hills, CA (1983)
4. Hall, E.T.: Beyond Culture. Doubleday, New York (1977)
5. Ting-Toomey, S., Gao, G., Trubisky, P., Yang, Z., Kim, H.S., Lin, S.L., Nishida, T.: Culture, face maintenance, and styles of handing interpersonal conflict: A study in five cultures. International Journal of Conflict Resolution 2, 275–296 (1991)
6. Ambady, N., Koo, J., Lee, F., Rosenthal, R.: More than words: Linguistic and nonlinguistic politeness in two cultures. Journal of Personality and Social Psychology 70, 996–1011 (1996)
7. Holtgraves, T.: Styles of language use: individual and cultural variability in conversational indirectness. Journal of Personality and Social Psychology 73, 624–637 (1997)
8. Hall, E.: The dance of life. Anchor Press, New York (1983)
9. Richmond, V.P., McCroskey, J.C.: Reliability and separation of factors on the assertiveness-responsiveness measure. Psychological Reports 67, 449–450 (1990)
10. Kume,T., Tokui, A., Hasegawa, N., Kodama, K.: A comparative study of communication styles among Japanese, Americans, and Chinese: Toward an Understand of Intercultural Friction. Retrieved in 2006 from http://coe-sun.kuis.ac.jp/public/paper/kuis/kume3.pdf
11. Gao, G., Ting-Toomey, S.: Communicating Effectively with the Chinese. Sage Publications, Thousand Oaks, CA (1998)
12. LeBaron, M.: Communication Tools for Understanding Cultural Differences. Retrieved in (2007), from http://www.beyondintractability.org/essay/communication_tools/

The Mobile Phone's Optimal Vibration Frequency in Mobile Environments

Jinho Yim[1], Rohae Myung[1], and Byongjun Lee[2]

[1] Department of Industrial Systems and Information Engineering, Korea University, 5-1 Anam-dong, Seongbuk-gu, Seoul, 136-785, Korea
[2] Department of Electrical Engineering, Korea University, 5-1 Anam-dong, Seongbuk-gu, Seoul, 136-785, Korea
{humans,rmyung,leeb}@korea.ac.kr

Abstract. Mobile environments are very dynamic and unpredictable [1]. When a mobile phone user is moving, his attention resources are reserved partly for passively monitoring and reacting to contexts and events and partly for actively constructing them [2]. In this paper, we suggest guidelines related to the optimal vibration frequency for the perception of mobile phone vibration when the user is moving. To guarantee the validity of this study, subjects were asked to indicate their perception of the randomly given 7 vibrotactile stimuli while they performed routine activities on a sidewalk, subway, or bus. With Logistic Regression analysis, the results showed that the optimal vibration frequency in the dynamic state was higher than 180 Hz, considerably higher than 151 Hz - the optimal vibration frequency obtained in the static state in the previous study. From this study, mobile phone manufacturers should consider this factor when designing the vibration frequency for the vibration mode so that missed calls in mobile environments are minimized.

Keywords: Multimodal, mobile environments, mobile phone, perception, optimal vibration frequency, missed call.

1 Introduction

Mobile phones offer multimodal feedback (such as visual, auditory, and vibration feedback) by essentially considering different usage environments so that users can set up a reception mode suitable to their situation in mobile environments. In particular, the vibration mode of a mobile phone allows its users to receive phone calls in noisy environments; the mode also serves to ensure propriety when the user is in a public place, even when the situation does not demand it. However, when people use a mobile phone in mobile environments, calls are frequently missed inadvertently.

It reported that a missed call results due to the reception mode settings and the carry mode of a mobile phone when the user is moving; these significantly influence the user's ability to perceive call reception. In addition, research on the general usage patterns of mobile phone users revealed that in mobile environments, phones were mostly set in the vibration mode and placed in trouser pockets, while the users

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 646–652, 2007.
© Springer-Verlag Berlin Heidelberg 2007

themselves walked or used some form of transportation. Moreover, users stated that they often missed calls when they were moving [3]. Here, the frequent use of the vibration mode can be attributed to the demands of modern life that require a person to be present in public places for long spells of time. Additionally, it can also be attributed to people not changing their call reception mode to the normal mode once they leave a public place.

In short, one of the reasons why users miss calls while moving is that they are unable to perceive the vibrations in the preset vibration mode. This is related to the vibrotactile perception sensitivity of the user and the carry mode of the user's mobile phone, as mentioned previously. In other words, in mobile environments, the phone could either be placed in the user's pocket, belt holder, or bag, or held in the user's hand. The vibrotactile perception sensitivity of the user diminishes owing to limited attentiveness and inadequate cognitive resources. In addition, there might also be a partial separation between the vibrotactile output device and the user's skin due to the presence of some material [4].

Currently, the frequency of the vibration motor in mobile phones is approximately between 130 Hz and 180 Hz (these results were obtained after the analysis of vibration motor specifications used by mobile phone manufacturing companies) with an average at 160 Hz (rotation speed: 10000 rpm). Nevertheless, missed calls do occur inadvertently, and therefore, we expect that a vibration frequency higher than the current default vibration frequency is required.

In this paper, we investigate the optimal vibration frequency for perception by a user in mobile environments. (Here, "mobile" implies a situation where a person or environment moves.) Our study results will provide basic research data for improvements in mobile phones in mobile environments.

2 Background

2.1 Limited Attentiveness to Mobile Phone in Mobile Environments

Mobile environments are very dynamic and unpredictable. When a mobile phone user is moving, his attention resources are reserved partly for passively monitoring and reacting to contexts and events and partly for actively constructing them [1, 2]. Therefore, since the user is moving and switching his attention according to the situation, he is unable to continuously pay attention only to the mobile phone.

As mentioned above, although people display limited attentiveness, their ability for vibrotactile perception enables immediate awareness, even if they are not extremely attentive. However, when people are moving, the vibrotactile perception threshold should be higher than that for the static state because of vibrations that occur spontaneously during movement. Accordingly, by appropriately designing the vibration mode, it should be possible to arrive at a suitable vibration frequency that considers this factor.

2.2 Vibrotactile Perception

Most of the literature available on vibrotactile perception focuses on direct contact with the skin (particularly the hand or the fingers) in the static state. Human skin is

very sensitive to vibrating stimuli at 230 Hz, regardless of the contact area, while it is insensitive to vibrations below approximately 100 Hz or above 600 Hz [5]. In other research on the absolute sensitivity of the hand toward vibrotactile stimulus, the frequency at which the hand was most sensitive was around 240 Hz, regardless of the contact area of stimulus and the region of the hand [6]. Further, it is showed that 120 Hz was the most effective frequency for transmitting information to the hand by vibrations [7]. The experiment using a handheld phone to determine the optimal vibration frequency of a mobile phone in the static state conducted that a frequency of 140~160 Hz (around 151 Hz) was sufficient to enable psychophysical recognition. In addition, it reported that a frequency of 151 Hz was more suitable than 120 Hz for discerning mobile phone vibrations [8].

However, in most actual usage environments, unlike the environments studied in the abovementioned researches, mobile phones are used in the dynamic state with indirect contact with the skin. Generally, a mobile phone is localized on a piece of cloth on the skin [9], and a space is formed between the phone and skin because people are usually standing. Moreover, in mobile environments, the user is constantly surrounded by noise made by vehicles, construction work, and so on, and street noise increases or decreases dynamically [3].

Thus, it can be explained that mobile phones are not always in direct contact with the skin, and therefore, the vibrotactile contact characteristics can vary according to the user's usage condition (such as standing, walking, and running).

The vibrotactile sensitivity varies on different parts of the body. It is important to consider the location of the phone with respect to the body when studying the vibrotactile interface since different locations have different levels of sensitivity and spatial acuity [10]. The skin on the fingertips and the lips is the most sensitive, while the leg is a relatively insensitive part of the body. Naturally, differences exist from person to person.

Thus, it can be explained that mobile phones are not always in direct contact with the skin, and therefore, the vibrotactile contact characteristics can vary according to the user's usage condition (such as standing, walking, and running).

3 Methods

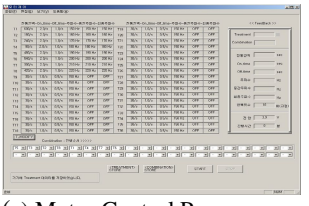
3.1 Subjects

Forty subjects (26 male, 14 female) participated in this experiment. All the participants were healthy and did not report any known neuropathologies that could affect their vibrotactile perception. The ages of the participants ranged between 22 and 45 years (with a mean age of 26.5 years).

3.2 Apparatus

For controlling the vibration frequency, we developed a program using C++ (Fig.1.a) and a prototype for vibration generation (hereafter referred to as "VibGen") that can operate in the stand-alone mode (Fig. 1.b). The phone used for the experiment was a general folding type model, and the portion where the vibration motor was located

(Fig. 1.c) was connected to the prototype. The vibration motor was a small coin-type DC motor. The frequency of the vibration motor ranged between 150 Hz and 250 Hz. The vibration frequency bandwidth was chosen because it was usable vibration motor specification that developed at present. However, mobile phone manufacturers usually do not use a vibration frequency greater than 217Hz. This is because of technical reasons such as occurrence of noise at 217Hz when TDMA (Time Division Multiple Access) is used.



(a) Motor Control Program.





(b) The VibGen.



(c) The test mobile phone

Fig. 1. Experiment in mobile environments

3.3 Experimental Design and Procedure

In this experiment, the independent variable was the frequency (ranging from 150 Hz to 210 Hz at intervals of 10 Hz) and environment (walking and mobile condition). The dependent variable was whether or not the vibrotactile stimulus was perceived. The seven frequencies consisted of 10 pulses (on time 2sec, off time 1 sec) were tested in random order, and given at irregular time intervals preset by the experimenter. The total experiment time was approximately 1h.

Currently experiment consists of two environment conditions: walking condition such as a sidewalk or public place, and mobile condition by vehicles such as subway or bus. Twenty subjects participated in each condition. The subjects placed the test mobile phone and the VibGen in each of their trouser front pockets and moved around in heavily or sparsely populated areas in the city (Fig. 2). All the actions of the subjects appeared extremely natural so that they did not pay particular attention to their mobile phones. Whenever the subjects sensed the suddenly transmitted vibrotactile stimulus, they responded to the experimenter by saying "Received." At this instant, the experimenter recorded on paper the time of response, subject behavior, and context.



Fig. 2. Experiment in mobile environments

4 Results

The ANOVA for the Vibrotactile perception, which had a significant effect on the ability to perceive reception while moving, was shown in Table 1. Since all interactions were not significant, main effect tests were meaningful. The Virotactile perception showed statistically significant differences in the vibration frequency ($p<0.000$). The Gender and Environment had no significant differences.

Table 1. ANOVA Table for the Vibrotactile perception

Source	df	SS	MS	F	Pr > F
Frequency	6	367.223	61.204	4.771	.000*
Gender	1	13.613	13.613	1.061	.304
Environment	1	27.611	27.611	2.153	.141
DB*N0	6	105.107	17.518	1.366	.230
DB*Gender	6	85.279	14.213	1.108	.359
Replication Error	210	2693.686	12.827		

* Significant at =0.05, R^2=0.525

In order to find the significant levels of frequency, SNK comparison test was conducted. The results were shown in Table. 2. and Fig. 3.

The results showed that the optimal vibration frequency, which had a significant effect on the ability to perceive reception while moving, was higher than 180Hz. Furthermore, the perception rate had a tendency to increase slightly beyond 210Hz. The result of the SNK test revealed no significant differences between the perception rates at 180 Hz and 210 Hz.

Table 2. SNK Results for the Vibrotactile perception

Frequency	N	Group		
		1	2	3
150Hz	33	1.61		
160Hz	33	1.73	1.73	
170Hz	33	2.03	2.03	
180Hz	33	3.94	3.94	3.94
190Hz	33	3.97	3.97	3.97
200Hz	33		4.15	4.15
210HZ	33			4.97
Significant probability		.60	.50	.647

* Significant at =0.05

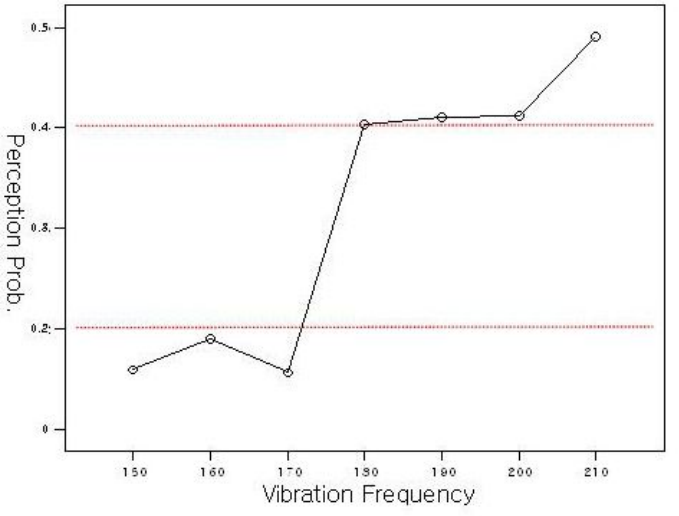


Fig. 3. Vibrotactile perception probability at each frequency

5 Discussions

In previous studies on vibrotactile perception, the optimal vibration frequency of a mobile phone was determined to be 151 Hz, which was sufficient for psychophysical recognition [8]. However, it was not applicable to actual usage environments because the experiments were conducted in the static state and with direct contact with the skin. Meanwhile, the result presented in this study can be identified as the optimal frequency for perception while moving. This is because; it has validity in that the experiment was conducted in the dynamic state in actual mobile environments and not in a laboratory.

The results showed that the optimal vibration frequency, which had a significant effect on the ability to perceive reception while moving, was higher than 180Hz. The

perception rate had a tendency to increase slightly beyond 210Hz. However the lower is a significant frequency in mobile phones as possible to save power consumption because the vibration motor was driven by voltages. Hence, even though statistically significant differences between the four frequencies were not observed, it is possible that 180 Hz is more appropriate as the optimal vibration frequency while moving.

6 Conclusions

This study was initiated because of a usage problem in mobile devices in mobile environments wherein calls are missed when users are unable to perceive the vibration in the preset vibration mode while they are moving. From this perspective, the results shown in this paper will have a serious impact on improving the vibration interface in the dynamic state.

The conclusions from this study can be summarized as follows.

1) The optimal vibration frequency for vibrotactile perception in mobile environment is found to be 180 Hz.
2) The vibrotactile perception threshold of a mobile phone appears to be higher in the dynamic state (180 Hz) than in the static state (151 Hz).

References

1. Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A.: Understanding Mobile Contexts. Personal and Ubiquitous Computing 8, 135–143 (2004)
2. Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J.: Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems, pp. 919–928 (2005)
3. Baek, Y., Myung, R., Yim, J.: A Study on Interaction between Multimodal Feedback Setting and Portable Patterns through Behavior Study of Mobile Phone User in Mobile Environments, HCI, Special Interest Groups of Human Computer Interface of Korea Information Science Society, vol. 1 (2006), pp. 579–586 (2006)
4. Kaaresoja, T., Linjama, J.: Perception of Short Tactile Pulses Generated by a Vibration Motor in a Mobile Phone, IEEE (2005)
5. Bliss, J.C., Katcher, M.H., Rogers, C.H., Shepard, R.P.: Optical-to Tactile Image Conversion for the Blind. IEEE Transactions on Man.-Machine Systems MMS-11, 58–64 (1974)
6. Lee, S.: Human Sensitivity Reponses to Vibrotactile Stimulation on the Hand: Measurement of Absolute Thresholds. Journal of Ergonomics Society of Korea 17(2), 1–10 (1998)
7. Lee, S.: Human Sensitivity Reponses to Vibrotactile Stimulation on the Hand: Measurement of Differential Thresholds. Journal of Ergonomics Society of Korea 18(3), 1–10 (1999)
8. Lee, B., Park, H., Myung, R.: JND-based mobile phone optimal vibration frequency. Journal of the Korean Institute of Industrial Engineers 30(1), 27–35 (2004)
9. Linjama, J., Puhakka, M., Kaaresoja, T.: User studies on tactile perception of vibrating alert. HCI International 2003 3, 280–284 (2003)
10. Brewster, S., Brown, L.M.: Tactons: Structured Tactile Messages for Non-Visual Information Display, Australasian User Interface Conference AUIC2004, pp. 15–23(2004)

A Comparative Study of Mid-market IT Customers in China and U.S.

Yi Ren Yuan[1] and Thomas Hogaboam[2]

[1] China Development Lab, IBM
9th Floor, Jin Mao Tower, 88 Century Boulevard, Pudong Shanghai, 2001211, PRC
yuanyr@cn.ibm.com
[2] Systems &Technology Group, IBM
IBM Rochester, 3605 Hwy 52 North, Rochester, MN 55901, United States
hogaboam@us.ibm.com

Abstract. Many companies are realizing the economic benefit of localizing products designs. Comparative studies are often used to obtain information needed to support these localized designs. Our experience with one such study revealed that it is also necessary to localize the study design itself. Additionally, results of this study indicate that there are significant differences in the structure and use of IT staff in the U.S. and China.

Keywords: Parallel study, SMB, Mid-market, Customer interview, Study design.

1 Background

China is expected to gain double digit IT growth over the next 8 years. It is a very important emerging market with its own specific characteristics. This market is different from U.S market in many ways, from government regulations to user cognitive models.

To provide more consumable product and services for this market, IBM formed a new User-Centered Design (UCD) team in CDL (China Development Lab) last year. The CDL UCD team focuses on driving localized solutions for China customers to develop better brand acceptance and emotional appeal in China market. This paper discusses one of the early studies undertaken by this new lab.

2 A Case Study in "Localizing the Study"

Small and median business (SMB) is a hot topic recently. SMB is the most active sector of the China economy, and represents the highest growth sector of the U.S. market for IT equipment. According to China government's data, ninety-nine percent of the Chinese companies are small and medium businesses. [1]

Because of this rapid growth, it is important for IT companies to understand the wants and needs of this market. During 2006, IBM conducted a comparative study of this market segment in both the U.S. and in China. The goal of this study was to

N. Aykin (Ed.): Usability and Internationalization, Part I, HCII 2007, LNCS 4559, pp. 653–657, 2007.
© Springer-Verlag Berlin Heidelberg 2007

obtain general information about IT staff characteristics, hardware and software inventories, company characteristics, allocation of IT staff effort, and IT pain points.

Originally, this study was designed for the U.S. market only. With the establishment of the CDL UCD lab in 2006 there was an opportunity to conduct a comparative study. This type of comparative study is crucial to the design of hardware and software products that are 'localized' for different geographies.

Our original plan was to use the same study design in both the U.S. and China. This is a typical approach to comparative studies because it allows direct comparisons of differing results. In statistical jargon, we planned to apply the same set of "treatments" to different "plots".

As the study unfolded in China, however, it became clear that modifications to the study design, procedures, and materials would be needed. While these modifications made subsequent interpretation of comparative data problematic in some cases, they also increased the amount of useful information collected in China. We refer to these design changes as "localizing the study". This remainder of this paper presents a case study of some of the differences in study design, procedures, materials, and results from this study.

2.1 Independent Variable Modifications

As is typical in many IT industry studies in the U.S., one of the independent variables was Industry Segment (manufacturing, wholesale, retail, etc). These segments are identified basing on common purchasing habits, IT needs and software needs, etc. However, there is no guarantee that the same market structure has much utility in different geographies such as China.

In particular, we found that it was difficult for customers to self-identify as either "wholesale" or "retail". There was no similar difficulty in the U.S. Rather than force an arbitrary classification on the China market we combined the wholesale and retail markets into one segment.

A related problem involves adjustments to sampling plans that call for proportional sampling. For instance, you might want the sample size from each market segment to be proportional to its size in the population. Because the population proportions differ in the U.S. and China, the sample proportions must also differ. So not only might the market segments be different, but even the sample size in comparable segments might be different.

A third problem we encountered involved the existence of important variables in one country that are known to play no role in other countries. In China, economic growth is geographically unbalanced. China uses different tiers to indicate the scale of the city and its economic development level. The tier 1 cities are biggest and most advanced cities like Beijing, Shanghai etc. Based on the assumption that economic growth unbalance might affect the IT operation model or IT maturity, we added "city tier" variable into the China's sampling plan.

2.2 Study Material Differences

The original design of the study in the U.S. called for three separate data collection effort for each customer that volunteered to participate:

1. A pre-visit questionnaire
2. A pre-visit conference call
3. A customer site visit

The pre-visit questionnaire was used to collect customers' background information and understand their basic IT environment. The pre-visit conference call was used to clarify and expand upon questionnaire answers, and the site visit was used to collect more extensive information about customer pain points and allocation of IT staff effort.

Of course, the first problem faced in administering an existing English questionnaire is translation. One interesting thing is Chinese IT persons' working language is a mixture of English and Chinese terms.For example, most Chinese engineers say High availability in Chinese. But most Chinese engineers say term Logical Partitioning as LPAR like U.S. engineers do, although this term has a standard Chinese translation. When translating the questionnaire, you must be careful about whether Chinese or English should be used. If it is hard to determine, then just keep both.

Also, you might need to change way how the question is asked to make it is more acceptable to the local customers. In our study, we would like to understand IT department's position in the whole company's structure. Instead of asking directly 'What is your manager's title?' it was better to ask 'who does your department report to?'. In China, "What is your manager's title" is a more personal question. Chinese customers usually are not willing answer the personal related questions when they are not familiar with you.

2.3 Procedural Modifications

As noted above, the original study plan called for three separate data collections efforts. It became quickly apparent that this approach was not going to work in China. Many Chinese customers were reluctant or simply refused to fill out the pre-visit survey, even after they had agreed to participate in the study. By way of contrast, this was not an issue in the U.S.

As a result of this, the Chinese customers were told that they did not need to answer any questions on the questionnaire that made them uncomfortable, and indeed the information collected by the pre-visit questionnaire was very incomplete.

The original plan called for completing any missing pre-visit questionnaire during a subsequent conference call. However, it also turned out that the Chinese customers were not accustomed to sharing this type of information on the telephone. The most common reply from early customers was "We can talk about it when you come to visit".

Consequently, we needed to collect most of the information from the pre-visit questionnaire during the actual site visit. This required adjustments to interviewing techniques and schedules. This went very well. The Chinese customers are more willing to talk and share their information in the face-to-face communication.

By way of contrast, most U.S. SMB customers did not want to devote a large amount of time to a site interview. They preferred the pre-visit questionnaire, which could be filled out at their convenience.

A final procedural modification involved a rearrangement of the sequence of questions. Questions that were viewed as 'more sensitive' by Chinese customers, such

as asking about their team structure, needed be placed at the end of the interview when the interviewer had built up sufficient rapport. While this same technique is used in any geography, the questions that are viewed as 'sensitive' are very different.

2.4 Differences in Results

There were many interesting differences between the U.S. and China SMB IT environments. Most U.S. mid-market companies have a longer history, having been in business for over 25 years. And most of the IT staff was very experienced, with most of them being between 40 and 60 years old. By contrast, China mid-market customers have been in business for about 12 years with an IT staff between 20 and 30 years old.

The China IT labor pool is not only younger, but more mobile. China mid-market companies report more problems with recruiting and retaining IT staff.

Given these differences, it should come as no surprise that China customers report more pain points that are related to procuring training for their IT staff.

There are additional IT staff differences worth noting. A typical mid-market company in China averages 11 staff members, while in the U.S. the average is 3. There is a corresponding difference in reliance on outside business partners for support of IT infrastructure. The average U.S. company works with 5 or 6 business partners, while the average Chinese company works with 3. Additionally, it is very common to find in-house software development skills in China, while this is very uncommon in the U.S. This fact accounts for some of the staff size difference.

These differences reflect a different approach to IT management. In the U.S. reduction in IT staff size has caused an increased reliance on external business partners, and an increased emphasis on the skills needed to coordinate and work with these business partners. Generically, these are project management skills.

In China we find a greater reliance on an in-house, "do it yourself" approach, with software development resources devoted to customizing third-party business applications to gain competitive advantage. This is an approach that is found in the U.S. for larger companies, but is not typical of many smaller companies. These differences are illustrated in Figure 1 and 2.

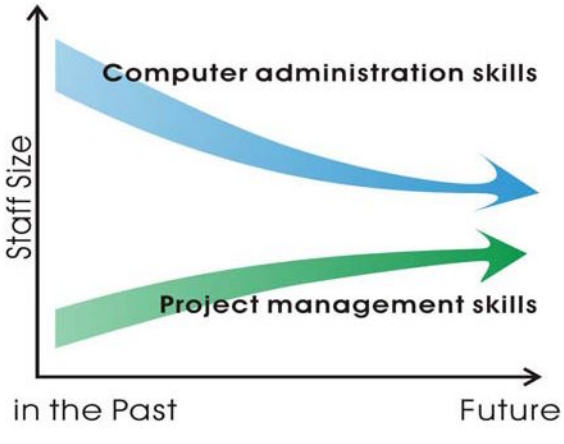


Fig. 1. U.S. customers' skill trend

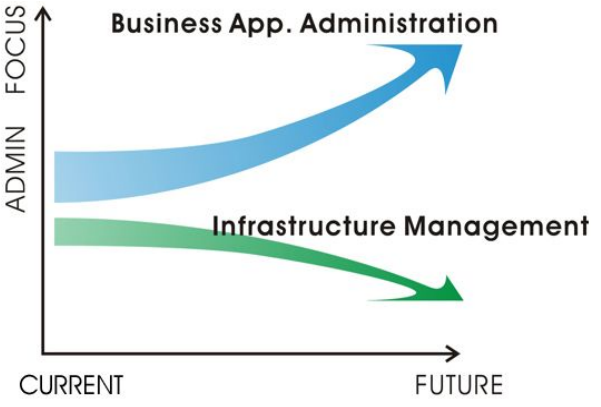


Fig. 2. Chinese customers' skill trend

3 Summary

This study of the U.S. and China SMB market illustrated some of the issues and problems involved in conducting comparative studies in different geographies. While seeking to uncover information needed to design and produce localized products, it becomes necessary to localize the study itself.

While it is often thought desirable to employ identical study designs in different geographies, difference in culture and market required modifications to our study design. In this study, modifications to the definitions of the independent variables, materials, and procedures were deemed necessary in order to collect comparable data.

Each modification, of course, introduces a possible source of difficulty in interpreting the resultant data. It is always possible that observed differenced in results are influenced by the design differences, so each change must be carefully considered.

Reference

1. http://www.sme.gov.cn/web/assembly/action/browsePage.do?channelID=10089&contentID=1167587742566

Author Index

Ajmera, Rahul 33
Alostath, Jasem M. 225
Anirudha, Joshi 108
Aykin, Nuray 3

Baker, Rebecca Matson Sukach 235
Batra, S. 243
Beekhuyzen, Jenine 586
Beelders, Tanya R. 250
Bishu, R.R. 243, 493
Blech, Michael 412
Blignaut, P.J. 250
Bødker, Mads 10
Braun, Kelly 291
Brevik, Eivind 258
Bygstad, Bendik 258

Chan, Alan H.S. 379
Chan, Yu-Ching 146
Chang, Ya-Ping 146
Chavan, Apala Lahiri 3, 19, 27, 33
Chavan, Sameer 174
Chen, Baihong 267
Chen, Chien-Hsiung 37
Chen, Kuohsiang 47
Chen, Shao-Nung 369
Chen, Vivian Hsueh-Hua 65
Chen, Zhongming 502
Cheng, Ricer 154
Chipchase, Jan 483
Chiu, Shu-chuan 47
Cho, Y.J. 631
Cho, Youngseok 405
Choi, Kueng Mi 550
Choi, Young Lim 550
Clemmensen, Torkil 274, 281, 317, 422, 462
Cui, Yanqing 483
Cui, Ming Hai 550

Dai, Guozhong 622
Darnell, Elissa 181
Dednam, E. 250
Diaz, Alvaro Enrique 57
Dong, Jianming 291
Dong, Ying 113
Douglas, Ian 297
Dray, Susan M. 3
Duda, Sabrina 606
Duh, Henry Been-Lirn 65

Egan, Richard 442

Fendler, Jens 452
Foucault, Brooke 74
Fujii, Kunikazu 186
Fujimura, Noriyuki 99

Ghinea, Gheorghita 258
Glucroft, Brian I. 83
Gnaneswaran, V. 493
González, María Paula 306
Granollers, Antoni 306
Griffiths, Lee 502

Hamasaki, Masahiro 99
Han, Sung H. 398, 405
Heimgärtner, Rüdiger 89
Hertzum, Morten 317
Hogaboam, Thomas 653
Hope, Tom 99
Hornbæk, Kasper 317
Hsieh, Yi-Chen 146
Huang, Chiwu 513
Huang, Yuan-Ching 146
Hwang, Jun-Lung 566
Hyun, H.J. 631

Ichikawa, Fumiko 483
Ishida, Keisuke 99

Ji, Yong Gu 327
Jin, Beomsuk 327
Jo, Jang Hyeon 541
Jung, Hyun-Wook 523

Kelkar, Kuldeep 291
Khalfan, Abdulwahed Moh 225
Kim, Byungjoo 113
Kim, Ji Hye 531
Kim, Jung-Yong 523
Kim, Sungjin 113

Kiris, Esin 235
Ko, Sangmin 327
Kuen Seong, Daniel Su 432
Kumar, Jyoti 281, 317, 336, 462
Kurniawan, Sri 596

Lee, Byongjun 646
Lee, Cheol 541, 559
Lee, Dong-Seok 346
Lee, Joo Hwan 541
Lee, Joohyun 550
Lee, Joong-Ho 130
Lee, Jungjoo 122
Lee, Kun-Pyo 113, 122, 531
Li, Christina 138
Li, Huiyang 281
Li, Jie 622
Liang, Sheau-Farn Max 355
Lim, J.S. 631
Lim, Ji Hyoun 559
Lin, Chiuhsiang Joe 566
Lin, Fang-chyuan 47
Lin, Rungtai 146, 154
Lisney, Eleanor 138
Liu, Chi-No 566
Liu, Sean 138
Lodge, Carol 365
Loréss, Jesús 306
Luh, Ding-Bang 369

Mannonen, Petri 388
McDonald, T. 250
Md Noor, Nor Laila 212
Mehad, Shafie 212
Melican, Jay 74
Milewski, Allen 442
Mun, Jaeseung 327
Mushtaha, Abdalghani 164
Myung, Rohae 646

Nakamura, Yoshiyuki 99
Nam, Yunja 550
Ng, Annie W.Y. 379
Nielsen, Janni 336
Nielsen, Lene 10, 174
Nieminen, Marko 576
Nieminen, Mika P. 388
Nishimura, Takuichi 99

Oikarinen, Anna 576
Orngreen, Rikke N. 10

Pan, Young-Hwan 346
Park, Sun Young 559
Park, J.S. 631
Park, Ji-Hyung 130
Park, Jungchul 398, 405
Park, K.R. 631
Park, Wonkyu 398
Park, Yong S. 398, 405
Peter, Christian 412
Plocher, Tom 274, 615
Prabhu, Girish 3

Qu, Weina 615

Rivera, Krisela 181
Ruth, Alison 586

Sa-nga-ngam, Prush 596
Schiessl, Michael 606
Schultz, Randolf 412
Shi, Qingxin 281, 317, 422
Shigenobu, Tomohiro 186
Siew Yen, Victoria Yee 432
Suh, Won Yong 541
Sun, Huatong 196
Sun, Ming-Xian 146, 154
Sun, Xianghong 281, 615
Sundling, Becky 206

Tao, Lin-Mi 472
Tran, Thu-Trang 122
Tremaine, Marilyn M. 442
Troyer, Olga De 164
Tsai, Chia-Ying 37
Tsai, Chieh-Ming 513

Urban, Bodo 412

Vasnaik, Omar 235
Viitanen, Johanna 388
Voskamp, Jörg 412

Wan Mohd Isa, Wan Abdul Rahim 212
Wang, Danli 622
Whang, Min Cheol 631
Winschiers, Heike 452

Xiao, Xiao-Ling 472
Xiao, Xingrong 637
Xu, Guang-You 472

Yammiyavar, Pradeep 281, 317, 336, 462
Yang, Huichul 398
Yang, Rong 267
Yeom, Ki-Won 130
Yim, Jinho 646
Yoshino, Takashi 186
You, Im Kyeong 531
Yuan, Yi Ren 653
Yun, Myung Hwan 541, 559

Zhang, Shaoke 637
Zhang, Suling 442
Zhang, Xiang 472
Zhao, Chen 637