Julie A. Jacko (Ed.)

LNCS 6761

Human-Computer Interaction

Design and Development Approaches

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

1 Part I

HCI2011 INTERNATIONAL

Springer

Lecture Notes in Computer Science 6761

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
Alfred Kobsa
University of California, Irvine, CA, 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
TU Dortmund University, Germany
Madhu Sudan
Microsoft Research, Cambridge, MA, USA
Demetri Terzopoulos
University of California, Los Angeles, CA, USA
Doug Tygar
University of California, Berkeley, CA, USA
Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

Julie A. Jacko (Ed.)

Human-Computer Interaction

Design and Development Approaches

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

Springer

Volume Editor

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

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21601-5 e-ISBN 978-3-642-21602-2
DOI 10.1007/978-3-642-21602-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929076

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

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

© Springer-Verlag Berlin Heidelberg 2011
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.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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

Printed on acid-free paper

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

Foreword

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

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

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

- HCI design
- Model-based and patterns-based design and development
- Cognitive, psychological and behavioral issues in HCI
- Development methods, algorithms, tools and environments
- Image processing and retrieval in HCI

The remaining volumes of the HCI International 2011 Proceedings are:

- Volume 2, LNCS 6762, Human–Computer Interaction—Interaction Techniques and Environments (Part II), edited by Julie A. Jacko
- Volume 3, LNCS 6763, Human–Computer Interaction—Towards Mobile and Intelligent Interaction Environments (Part III), edited by Julie A. Jacko
- Volume 4, LNCS 6764, Human–Computer Interaction—Users and Applications (Part IV), edited by Julie A. Jacko
- Volume 5, LNCS 6765, Universal Access in Human–Computer Interaction—Design for All and eInclusion (Part I), edited by Constantine Stephanidis
- Volume 6, LNCS 6766, Universal Access in Human–Computer Interaction—Users Diversity (Part II), edited by Constantine Stephanidis
- Volume 7, LNCS 6767, Universal Access in Human–Computer Interaction—Context Diversity (Part III), edited by Constantine Stephanidis

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

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

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

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

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

July 2011 Constantine Stephanidis

Organization

Ergonomics and Health Aspects of Work with Computers

Program Chair: Michelle M. Robertson

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

Human Interface and the Management of Information

Program Chair: Michael J. Smith

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

Human–Computer Interaction

Program Chair: Julie A. Jacko

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

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

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

Universal Access in Human–Computer Interaction

Program Chair: Constantine Stephanidis

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

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

Virtual and Mixed Reality

Program Chair: Randall Shumaker

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

Internationalization, Design and Global Development

Program Chair: P.L. Patrick Rau

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

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, Panayiotis Zaphiris

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

Augmented Cognition

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

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

Digital Human Modeling

Program Chair: Vincent G. Duffy

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

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

Human-Centered Design

Program Chair: Masaaki Kurosu

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

Design, User Experience, and Usability

Program Chair: Aaron Marcus

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

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

HCI International 2013

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

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

Table of Contents – Part I

Part I: HCI International 2011 Keynote Speech

Technology-Mediated Social Participation: The Next 25 Years of HCI Challenges 3
Ben Shneiderman

Part II: HCI Design

Towards a Cognitive-Based User Interface Design Framework Development 17
Natrah Abdullah, Wan Adilah Wan Adnan, and Nor Laila Md Noor

A Design Science Framework for Designing and Assessing User Experience 25
Sisira Adikari, Craig McDonald, and John Campbell

Objective and Subjective Measures of Visual Aesthetics of Website Interface Design: The Two Sides of the Coin 35
Ahamed Altaboli and Yingzi Lin

Interaction Design Teaching Method Design 45
Chen-Wei Chiang and Kiyoshi Tomimatsu

Designing Interaction Concepts, Managing Customer Expectation and Mastering Agile Development in Rich Application Product Development 54
Marcela Esteves and Vladimir Andrade

POLVO - Software for Prototyping of Low-Fidelity Interfaces in Agile Development 63
Júnia Gonçalves and Caroline Santos

Towards a Conceptual Framework for Interaction Design for the Pragmatic Web 72
Heiko Hornung and M. Cecília C. Baranauskas

Blowing Light: Green-Based Interaction Design 82
Yu-Chun Annester Huang, Chih-Chieh Tsai, Teng-Wen Chang, Pen-Yan Tsai, Tien-Hsin Hung, and Jai-Jung Chen

The Impact of Rich Application Frameworks on User Experience Design 92
Tobias Komischke

Axiomatic Design for Biometric Icons 98
Sheau-Farn Max Liang and Chien-Tsen Lin

A Rapid Prototyping Tool for Interactive Device Development 107
Mark Merlo and Mark Bachman

Key Requirements for Integrating Usability Engineering and Software Engineering 114
Karsten Nebe and Volker Paelke

Message Oriented Middleware for Flexible Wizard of Oz Experiments in HCI 121
Mirko Otto, Rafael Friesen, and Dietmar Rösner

Design and Rich Application Frameworks 131
Kevin H. Richardson

Enhancing Personas with Their Main Scenarios 136
Alícia Valls, Muriel Garreta-Domingo, and Marta López

Streamlining User Experience Design and Development: Roles, Tasks and Workflow of Applying Rich Application Technologies 142
Xianjun Sam Zheng, Mo Wang, Gilberto Matos, and Shaopeng Zhang

Part III: Model-Based and Patterns-Based Design and Development

Configurable Executable Task Models Supporting the Transition from Design Time to Runtime 155
Birgit Bomsdorf, Stefan Grau, Martin Hudasch, and Jan-Torsten Milde

Automatic Adaptation of User Workflows within Model-Based User Interface Generation during Runtime on the Example of the SmartMote 165
Kai Breiner, Kai Bizik, Thilo Rauch, Marc Seissler, Gerrit Meixner, and Philipp Diebold

Towards an Automatic Analysis of Interaction Data for HCI Evaluation: Application to a Transport Network Supervision System 175
Selem Charfi, Houcine Ezzedine, Christophe Kolski, and Faouzi Moussa

A Formal Model of Mixed-Initiative Interaction in Design Exploration 185
Sambit Datta and Michael Hobbs

Intertwined Modeling and Implementation of Interactive Systems Using HOPS 194
Anke Dittmar and Peter Forbrig

HCI Patterns as a Means to Transform Interactive User Interfaces to Diverse Contexts of Use 204
Jürgen Engel, Christian Märtin, and Peter Forbrig

Process Choreography for Human Interaction Computer-Aided Simulation 214
Carlos Fernandez-Llatas, Juan Bautista Mocholí, Pilar Sala, and Juan Carlos Naranjo

Building Multimodal Interfaces Out of Executable, Model-Based Interactors and Mappings 221
Sebastian Feuerstack and Ednaldo Pizzolato

The First Interaction Design Pattern Library for Internet of Things User Created Applications 229
Marc Godon, Mohamed Ali Feki, Marc Roelands, and Lieven Trappeniers

Differentiating between Successful and Less Successful Products by Using MAInEEAC – A Model for Interaction Characterization 238
Steffen Hess, Andreas Maier, and Marcus Trapp

Patterns for *Usable* Accessible Design 248
Homa Javahery, Michael Gower, Daniel Sinnig, and Peter Forbrig

From Structural Analysis to Scenarios and Patterns for Knowledge Sharing Applications 258
Claus Kaelber and Christian Märtin

A Design Patterns Approach to Adaptive User Interfaces for Users with Special Needs 268
Matthias Peissner, Andreas Schuller, and Dieter Spath

User Interface Representation Using Simple Components 278
Javier Rodeiro Iglesias and Pedro M. Teixeira-Faria

Model-Based Ubiquitous Interaction Concepts and Contexts in Public Systems 288
Thomas Schlegel and Christine Keller

Towards Pattern-Driven Engineering of Run-Time Adaptive User Interfaces for Smart Production Environments 299
Marc Seissler, Kai Breiner, and Gerrit Meixner

Complex Components Abstraction in Graphical User Interfaces 309
Pedro M. Teixeira-Faria and Javier Rodeiro Iglesias

User-Oriented Accessibility Patterns for Smart Environments 319
Michael Zaki and Peter Forbrig

Part IV: Cognitive, Psychological and Behavioural Issues in HCI

Visual Hierarchy and Viewing Behavior: An Eye Tracking Study 331
Soussan Djamasbi, Marisa Siegel, and Tom Tullis

Cognitive Analysis for Knowledge Modeling in Air Traffic Control Work 341
Satoru Inoue, Hisae Aoyama, and Keiichi Nakata

Individual Differences in Work Load While Doing Multitasking with a Computer 351
Kari Kallinen, Inger Ekman, and Niklas Ravaja

Finding a Relationship between Internet Anxiety and Human Behavior 359
Santosh Kumar Kalwar, Kari Heikkinen, and Jari Porras

Human Behavioral Simulation Using Affordance-Based Agent Model ... 368
Namhun Kim, Jaekoo Joo, Ling Rothrock, Richard Wysk, and Young-Jun Son

Investigating the Effects of Metacognition in Dynamic Control Tasks ... 378
Jung Hyup Kim, Ling Rothrock, Anand Tharanathan, and Hari Thiruvengada

The Effects of Personality Type in User-Centered Appraisal Systems ... 388
Zacharias Lekkas, Nikos Tsianos, Panagiotis Germanakos, Constantinos Mourlas, and George Samaras

Measuring Corrective Reaction Time with the Intermittent Illumination Model 397
Jui-Feng Lin, Colin G. Drury, Chin-Mei Chou, Yu-De Lin, and Yi-Quan Lin

Psychometric Evaluation with Brain-Computer Interface 406
Paolo Perego, Anna Carla Turconi, Chiara Gagliardi, and Giuseppe Andreoni

An Inductive Inference Model to Elicit Noncompensatory Judgment Strategies 414
Jing Yin and Ling Rothrock

A User-Friendly Tool for Detecting the Stress Level in a Person's Daily Life 423
Irene Zaragozá, Beatriz Rey, Cristina Botella, Rosa Baños, Inés Moragrega, Diana Castilla, and Mariano Alcañiz

"How Do I Line Up?": Reducing Mental Transformations to Improve Performance 432
Guy W. Zimmerman, Dale Klopfer, G. Michael Poor, Julie Barnes, Laura Leventhal, and Samuel D. Jaffee

Part V: Development Methods, Algorithms, Tools and Environments

A Middleware Architecture for Designing TV-Based Adapted Applications for the Elderly 443
Carlos Gacimartín, José Alberto Hernández, and David Larrabeiti

Performance Visualization for Large-Scale Computing Systems: A Literature Review 450
Qin Gao, Xuhui Zhang, Pei-Luen Patrick Rau, Anthony A. Maciejewski, and Howard Jay Siegel

Developing a User Recommendation Engine on Twitter Using Estimated Latent Topics 461
Hiroyuki Koga and Tadahiro Taniguchi

Project and Development of ErgoCoIn Version 2.0 471
Marcelo Morandini, Roberto Leite de Moraes Rodrigues, Marcus Vinicius Cerrato, and Marcos Lordello Chaim

A Reference Model for Adaptive Visualization Systems 480
Kawa Nazemi, Christian Stab, and Arjan Kuijper

A Proposal of an Interactive Music Composition System Using Gibbs Sampler 490
Akira Shirai and Tadahiro Taniguchi

Sensing User Needs: Recognition Technologies and User Models for Adaptive User Interfaces 498
Barnabas Takacs, Lajos Simon, and Matthias Peissner

uPlatform: A Customizable Multi-user Windowing System for Interactive Tabletop 507
Chenjun Wu, Yue Suo, Chun Yu, Yuanchun Shi, and Yongqiang Qin

Synchronization and Fluctuation of Rhythm in Musical Cooperative Performance 517
Tomohito Yamamoto

GMM Parameter Estimation by Means of EM and Genetic Algorithms 527
Sergey Zablotskiy, Teerat Pitakrat, Kseniya Zablotskaya, and Wolfgang Minker

Part VI: Image Processing and Retrieval in HCI

Shape – Based Human Actions Recognition in Videos 539
Nitish Amraji, Lin Mu, and Mariofanna Milanova

Dynamic Queries with Relevance Feedback for Content Based Image Retrieval 547
Murat Birinci, Esin Guldogan, and Moncef Gabbouj

Face Sketch Synthesis via Multivariate Output Regression 555
Liang Chang, Mingquan Zhou, Xiaoming Deng, Zhongke Wu, and Yanjun Han

Experimental Studies of Visual Models in Automatic Image Annotation 562
Ping Guo, Tao Wan, and Jin Ma

An Image Segmentation Method for Chinese Paintings by Combining Deformable Models with Graph Cuts 571
Ning He and Ke Lu

Image Skeletonization Based on Curve Skeleton Extraction 580
Xiaoyan Hu, Bo Sun, Huiqin Zhao, Bin Xie, and Hao Wu

Appearance Similarity Index for Medicinal Ampoule Labels 588
Masaomi Kimura, Yutaroh Furukawa, Akira Kojo, Hirotsugu Ishida, Keita Nabeta, Michiko Ohkura, and Fumito Tsuchiya

A Novel Parallel Clustering Algorithm Based on Artificial Immune Network Using nVidia CUDA Framework 598
Ruiyi Luo and Qian Yin

A Detection Method of Basic Mouth Shapes from Japanese Utterance Images 608
Tsuyoshi Miyazaki, Toyoshiro Nakashima, and Naohiro Ishii

Improving the Usability of Hierarchical Representations for Interactively Labeling Large Image Data Sets 618
Julia Moehrmann, Stefan Bernstein, Thomas Schlegel, Günter Werner, and Gunther Heidemann

The Similarity Index of Character Shape of Medicine Names Based on Character Shape Similarity (II) 628
Keita Nabeta, Akira Hatano, Hirotsugu Ishida, Masaomi Kimura, Michiko Ohkura, and Fumito Tsuchiya

ColoriT: Color Based Image Code Application to Aid in Memory Restoration of Offline Photo Artifacts 637
James Park, Jonghoon Seo, Ji-Hye Choi, and Tackdon Han

A Configurable Photo Browser Framework for Large Image Collections 643
Frode Eika Sandnes

Imaged Based Codes Performance Comparison for Mobile Environments 653
Jonghoon Seo, Ji Hye Choi, and Tack-don Han

A Global Optimal Algorithm for Camera Calibration with One-Dimensional Objects 660
Liang Wang, FuQing Duan, and Chao Liang

LSCM Based Non-rigid Registration for Craniofacial Surfaces 670
Wenkui Xie, Fuqing Duan, Qingqiong Deng, Mingquan Zhou, Zhongke Wu, and Liang Chang

High-Quality Fast Image Upsampling Algorithm Based on CUDA 677
Qingqing Xu, Xin Zheng, and Jie Chen

A Cubic Polynomial Model for Fisheye Camera 684
Haijiaing Zhu, Xiupu Yin, and Jinglin Zhou

Author Index 695

Table of Contents – Part II

Part I: Touch-Based and Haptic Interaction

Development of a High Definition Haptic Rendering for Stability and Fidelity 3
Katsuhito Akahane, Takeo Hamada, Takehiko Yamaguchi, and Makoto Sato

Designing a Better Morning: A Study on Large Scale Touch Interface Design 13
Onur Asan, Mark Omernick, Dain Peer, and Enid Montague

Experimental Evaluations of Touch Interaction Considering Automotive Requirements 23
Andreas Haslbeck, Severina Popova, Michael Krause, Katrina Pecot, Jürgen Mayer, and Klaus Bengler

More Than Speed? An Empirical Study of Touchscreens and Body Awareness on an Object Manipulation Task 33
Rachelle Kristof Hippler, Dale S. Klopfer, Laura Marie Leventhal, G. Michael Poor, Brandi A. Klein, and Samuel D. Jaffee

TiMBA – Tangible User Interface for Model Building and Analysis 43
Chih-Pin Hsiao and Brian R. Johnson

Musical Skin: A Dynamic Interface for Musical Performance 53
Heng Jiang, Teng-Wen Chang, and Cha-Lin Liu

Analyzing User Behavior within a Haptic System 62
Steve Johnson, Yueqing Li, Chang Soo Nam, and Takehiko Yamaguchi

Usability Testing of the Interaction of Novices with a Multi-touch Table in Semi Public Space 71
Markus Jokisch, Thomas Bartoschek, and Angela Schwering

Niboshi for Slate Devices: A Japanese Input Method Using Multi-touch for Slate Devices 81
Gimpei Kimioka, Buntarou Shizuki, and Jiro Tanaka

An Investigation on Requirements for Co-located Group-Work Using Multitouch-, Pen-Based- and Tangible-Interaction 90
Karsten Nebe, Tobias Müller, and Florian Klompmaker

Exploiting New Interaction Techniques for Disaster Control Management Using Multitouch-, Tangible- and Pen-Based-Interaction 100
Karsten Nebe, Florian Klompmaker, Helge Jung, and Holger Fischer

Saving and Restoring Mechanisms for Tangible User Interfaces through Tangible Active Objects 110
Eckard Riedenklau, Thomas Hermann, and Helge Ritter

Needle Insertion Simulator with Haptic Feedback 119
Seungjae Shin, Wanjoo Park, Hyunchul Cho, Sehyung Park, and Laehyun Kim

Measurement of Driver's Distraction for an Early Prove of Concepts in Automotive Industry at the Example of the Development of a Haptic Touchpad 125
Roland Spies, Andreas Blattner, Christian Lange, Martin Wohlfarter, Klaus Bengler, and Werner Hamberger

A Tabletop-Based Real-World-Oriented Interface 133
Hiroshi Takeda, Hidetoshi Miyao, Minoru Maruyama, and David Asano

What You Feel Is What I Do: A Study of Dynamic Haptic Interaction in Distributed Collaborative Virtual Environment 140
Sehat Ullah, Xianging Liu, Samir Otmane, Paul Richard, and Malik Mallem

A Framework Interweaving Tangible Objects, Surfaces and Spaces 148
Andy Wu, Jayraj Jog, Sam Mendenhall, and Ali Mazalek

The Effect of Haptic Cues on Working Memory in 3D Menu Selection 158
Takehiko Yamaguchi, Damien Chamaret, and Paul Richard

Part II: Gaze and Gesture-Based Interaction

Face Recognition Using Local Graph Structure (LGS) 169
Eimad E.A. Abusham and Housam K. Bashir

Eye-gaze Detection by Image Analysis under Natural Light 176
Kiyohiko Abe, Shoichi Ohi, and Minoru Ohyama

Multi-user Pointing and Gesture Interaction for Large Screen Using Infrared Emitters and Accelerometers 185
Leonardo Angelini, Maurizio Caon, Stefano Carrino, Omar Abou Khaled, and Elena Mugellini

Gesture Identification Based on Zone Entry and Axis Crossing 194
Ryosuke Aoki, Yutaka Karatsu, Masayuki Ihara, Atsuhiko Maeda, Minoru Kobayashi, and Shingo Kagami

Attentive User Interface for Interaction within Virtual Reality Environments Based on Gaze Analysis 204
Florin Barbuceanu, Csaba Antonya, Mihai Duguleana, and Zoltan Rusak

A Low-Cost Natural User Interaction Based on a Camera Hand-Gestures Recognizer 214
Mohamed-Ikbel Boulabiar, Thomas Burger, Franck Poirier, and Gilles Coppin

Head-Computer Interface: A Multimodal Approach to Navigate through Real and Virtual Worlds 222
Francesco Carrino, Julien Tscherrig, Elena Mugellini, Omar Abou Khaled, and Rolf Ingold

3D-Position Estimation for Hand Gesture Interface Using a Single Camera 231
Seung-Hwan Choi, Ji-Hyeong Han, and Jong-Hwan Kim

Hand Gesture for Taking Self Portrait 238
Shaowei Chu and Jiro Tanaka

Hidden-Markov-Model-Based Hand Gesture Recognition Techniques Used for a Human-Robot Interaction System 248
Chin-Shyurng Fahn and Keng-Yu Chu

Manual and Accelerometer Analysis of Head Nodding Patterns in Goal-oriented Dialogues.. 259
Masashi Inoue, Toshio Irino, Nobuhiro Furuyama, Ryoko Hanada, Takako Ichinomiya, and Hiroyasu Massaki

Facial Expression Recognition Using AAMICPF 268
Jun-Sung Lee, Chi-Min Oh, and Chil-Woo Lee

Verification of Two Models of Ballistic Movements 275
Jui-Feng Lin and Colin G. Drury

Gesture Based Automating Household Appliances 285
Wei Lun Ng, Chee Kyun Ng, Nor Kamariah Noordin, and Borhanuddin Mohd. Ali

Upper Body Gesture Recognition for Human-Robot Interaction 294
Chi-Min Oh, Md. Zahidul Islam, Jun-Sung Lee, Chil-Woo Lee, and In-So Kweon

Gaze-Directed Hands-Free Interface for Mobile Interaction 304
Gie-seo Park, Jong-gil Ahn, and Gerard J. Kim

Eye-Movement-Based Instantaneous Cognition Model for Non-verbal Smooth Closed Figures 314
Yuzo Takahashi and Shoko Koshi

Part III: Voice, Natural Language and Dialogue

VOSS -A Voice Operated Suite for the Barbadian Vernacular 325
David Byer and Colin Depradine

New Techniques for Merging Text Versions 331
Darius Dadgari and Wolfgang Stuerzlinger

Modeling the Rhetoric of Human-Computer Interaction 341
Iris Howley and Carolyn Penstein Rosé

Recommendation System Based on Interaction with Multiple Agents for Users with Vague Intention 351
Itaru Kuramoto, Atsushi Yasuda, Mitsuru Minakuchi, and Yoshihiro Tsujino

A Review of Personality in Voice-Based Man Machine Interaction 358
Florian Metze, Alan Black, and Tim Polzehl

Can Indicating Translation Accuracy Encourage People to Rectify Inaccurate Translations? 368
Mai Miyabe and Takashi Yoshino

Design of a Face-to-Face Multilingual Communication System for a Handheld Device in the Medical Field 378
Shun Ozaki, Takuo Matsunobe, Takashi Yoshino, and Aguri Shigeno

Computer Assistance in Bilingual Task-Oriented Human-Human Dialogues 387
Sven Schmeier, Matthias Rebel, and Renlong Ai

Developing and Exploiting a Multilingual Grammar for Human-Computer Interaction 396
Xian Zhang, Rico Andrich, and Dietmar Rösner

Part IV: Novel Interaction Techniques and Devices

Dancing Skin: An Interactive Device for Motion 409
Sheng-Han Chen, Teng-Wen Chang, and Sheng-Cheng Shih

A Hybrid Brain-Computer Interface for Smart Home Control 417
Günter Edlinger, Clemens Holzner, and Christoph Guger

Integrated Context-Aware and Cloud-Based Adaptive Home Screens for Android Phones 427
Tor-Morten Grønli, Jarle Hansen, and Gheorghita Ghinea

Evaluation of User Support of a Hemispherical Sub-Display with GUI Pointing Functions 436
Shinichi Ike, Saya Yokoyama, Yuya Yamanishi, Naohisa Matsuuchi, Kazunori Shimamura, Takumi Yamaguchi, and Haruya Shiba

Uni-model Human System Interface Using sEMG 446
Srinivasan Jayaraman and Venkatesh Balasubramanian

An Assistive Bi-modal User Interface Integrating Multi-channel Speech Recognition and Computer Vision 454
Alexey Karpov, Andrey Ronzhin, and Irina Kipyatkova

A Method of Multiple Odors Detection and Recognition 464
Dong-Kyu Kim, Yong-Wan Roh, and Kwang-Seok Hong

Report on a Preliminary Study Using Breath Control and a Virtual Jogging Scenario as Biofeedback for Resilience Training 474
Jacquelyn Ford Morie, Eric Chance, and J. Galen Buckwalter

Low Power Wireless EEG Headset for BCI Applications 481
Shrishail Patki, Bernard Grundlehner, Toru Nakada, and Julien Penders

Virtual Mouse: A Low Cost Proximity-Based Gestural Pointing Device 491
Sheng Kai Tang, Wen Chieh Tseng, Wei Wen Luo, Kuo Chung Chiu, Sheng Ta Lin, and Yen Ping Liu

Innovative User Interfaces for Wearable Computers in Real Augmented Environment 500
Yun Zhou, Bertrand David, and René Chalon

Part V: Avatars and Embodied Interaction

Influence of Prior Knowledge and Embodiment on Human-Agent Interaction 513
Yugo Hayashi, Victor V. Kryssanov, Kazuhisa Miwa, and Hitoshi Ogawa

The Effect of Physical Embodiment of an Animal Robot on Affective Prosody Recognition 523
Myounghoon Jeon and Infantdani A. Rayan

Older User-Computer Interaction on the Internet: How Conversational Agents Can Help 533
Wi-Suk Kwon, Veena Chattaraman, Soo In Shim, Hanan Alnizami, and Juan Gilbert

An Avatar-Based Help System for Web-Portals 537
Helmut Lang, Christian Mosch, Bastian Boegel, David Michel Benoit, and Wolfgang Minker

mediRobbi: An Interactive Companion for Pediatric Patients during Hospital Visit 547
Szu-Chia Lu, Nicole Blackwell, and Ellen Yi-Luen Do

Design of Shadows on the OHP Metaphor-Based Presentation Interface Which Visualizes a Presenter's Actions 557
Yuichi Murata, Kazutaka Kurihara, Toshio Mochizuki, Buntarou Shizuki, and Jiro Tanaka

Web-Based Nonverbal Communication Interface Using 3DAgents with Natural Gestures 565
Toshiya Naka and Toru Ishida

Taking Turns in Flying with a Virtual Wingman 575
Pim Nauts, Willem van Doesburg, Emiel Krahmer, and Anita Cremers

A Configuration Method of Visual Media by Using Characters of Audiences for Embodied Sport Cheering 585
Kentaro Okamoto, Michiya Yamamoto, and Tomio Watanabe

Introducing Animatronics to HCI: Extending Reality-Based Interaction 593
G. Michael Poor and Robert J.K. Jacob

Development of Embodied Visual Effects Which Expand the Presentation Motion of Emphasis and Indication 603
Yuya Takao, Michiya Yamamoto, and Tomio Watanabe

Experimental Study on Appropriate Reality of Agents as a Multi-modal Interface for Human-Computer Interaction 613
Kaori Tanaka, Tatsunori Matsui, and Kazuaki Kojima

Author Index 623

Table of Contents – Part III

Part I: Mobile Interaction

Field to File: A Tool for Activity Documentation Work in Remote Mobility Environments 3
Raúl Casillas and Alberto L. Morán

Trends, Challenges and Promises of Rich Experience on Mobile Devices 13
Yihsiu Chen

Finding Suitable Candidates: The Design of a Mobile Volunteering Matching System 21
Wei-Chia Chen, Yun-Maw Cheng, Frode Eika Sandnes, and Chao-Lung Lee

The Effort of Social Networking on Social Behavior – Integrating Twitter, Mobile Devices, and Wearable Clothing as an Example 30
Chen-Wei Chiang and Kiyoshi Tomimatsu

Computer Support of Team Work on Mobile Devices 38
Hilko Donker and Malte Blumberg

ProJest: Enabling Higher Levels of Collaboration Using Today's Mobile Devices 48
Babak Forutanpour and Jianfeng Ren

The Effect of Time Orientation and Representation of Points of Interests on the Use of Mobile Tour Guide 59
Fei Gao and Qin Gao

The Virtual Workplace of a Mobile Employee – How Does Vischer's Model Function in Identifying Physical, Functional and Psychosocial Fit? 69
Ursula Hyrkkänen and Suvi Nenonen

CornerPen: Smart Phone is the Pen 76
Bong-gyu Jang, Myonghee Lee, and Gerard J. Kim

Evaluation of Continuous Practice by Mobile Learning in Nursing Practical Training 84
Yukie Majima, Yumiko Nakamura, Yasuko Maekawa, Mizuko Hiramatsu, Yukari Nakajima, Satoshi Horii, and Hifumi Aoyama

XML in Formal Specification, Verification and Generation of Mobile HCI 92
Ines Riahi, Meriem Riahi, and Faouzi Moussa

An Efficient Document Browsing Method with Floating Diagram Window on Mobile Device 101
Yu Shibuya, Kazunobu Nagata, and Kazuyoshi Murata

Mobile Reminder for Flexible and Safe Medication Schedule for Home Users 107
Pei-Hsuan Tsai, Chi-Sheng Shih, and Jane W.-S. Liu

Enabling Efficient Browsing and Manipulation of Web Tables on Smartphone 117
Wenchang Xu and Yuanchun Shi

Part II: Interaction in Intelligent Environments

User Interface Framework for Ambient Intelligence Platforms 129
Patricia Abril-Jiménez, Cecilia Vera-Muñoz, María Teresa Arredondo Waldmeyer, Haydee Alvarez, and José R. Baragaño Galán

Scratchable Devices: User-Friendly Programming for Household Appliances 137
Jordan Ash, Monica Babes, Gal Cohen, Sameen Jalal, Sam Lichtenberg, Michael Littman, Vukosi Marivate, Phillip Quiza, Blase Ur, and Emily Zhang

Passive Identification and Control of Arbitrary Devices in Smart Environments 147
Andreas Braun and Felix Kamieth

Studying the Role of Interactivity in Museums: Designing and Comparing Multimedia Installations 155
Pedro Campos, Miguel Campos, João Pestana, and Joaquim Jorge

ARAMIS: Toward a Hybrid Approach for Human-Environment Interaction 165
Stefano Carrino, Elena Mugellini, Omar Abou Khaled, and Rolf Ingold

Express Yourself: Designing Interactive Products with Implicitness to Improve Social Interaction 175
Huang-Ming Chang and Rung-Huei Liang

Mojo iCuisine: The Design and Implementation of an Interactive Restaurant Tabletop Menu 185
Ting-Han Chen, Hsin-Hou Lin, and Yi-Di Yen

Usability of Nomadic User Interfaces 195
Walter Dees

Adaptive Implicit Interaction for Healthy Nutrition and Food Intake Supervision 205
Felix Kamieth, Andreas Braun, and Christian Schlehuber

Recall and Communication Support System for Reminiscences Triggered by Humming 213
Yusuke Kita and Yoshio Nakatani

Research of Passive Mode Interaction in Pervasive Computing 220
Yin Lu, Kejian Miao, Zhanhuai Li, and Ke Wei

Activity Recognition for Risk Management with Installed Sensor in Smart and Cell Phone 230
Daisuke Honda, Nobuchika Sakata, and Shogo Nishida

Can Twitter Be an Alternative of Real-World Sensors? 240
Tetsuro Takahashi, Shuya Abe, and Nobuyuki Igata

Reacting with Care: The Hybrid Interaction Types in a Sensible Space 250
Guo-Jhen Yu and Teng-Wen Chang

GoCoBa: Interactive Installation Design Applied on Combination of Context and People 259
Jia-Xuan Zhan and Kuo-Kuang Fan

Part III: Orientation and Navigation

Behavioral Cost-Based Recommendation Model for Wanderers in Town 271
Kenro Aihara, Hitoshi Koshiba, and Hideaki Takeda

A Framework for Agent-Based Simulation in Tourism Planning 280
Dingding Chao, Kazuo Furuta, and Taro Kanno

Safe-in-Place Awareness GPS System with Distance-Based and Duration-Based Notification Control 288
Chi Nung Chu and Gene Chu

Landmarks Detection to Assist the Navigation of Visually Impaired People 293
Paulo Costa, Hugo Fernandes, Verónica Vasconcelos, Paulo Coelho, João Barroso, and Leontios Hadjileontiadis

Interaction in Mobility: The Evaluation of Interactive Systems Used by Travellers in Transportation Contexts 301
Christophe Kolski, Guillaume Uster, Jean-Marc Robert, Kathia Oliveira, and Bertrand David

Evaluation of Wayfinding Performance and Workload on Electronic Map Interface 311
Ya-Li Lin and Cheng-Han Wang

Implementing Effective Tactile Symbology for Orientation and Navigation 321
Bruce Mortimer, Gary Zets, Greg Mort, and Curtis Shovan

Using Sound Patterns to Enhance Directional Sound for Emergency Route Guidance 329
Tom Plocher, Zhaoxia Janet Jin, and Foong-Yeen Donny Chan

A Knowledge Elicitation Study for a Speech Enabled GIS to Handle Vagueness in Communication 338
Hongmei Wang

Believe What You Hear, Not What You See – Vision Interferes with Auditory Route Guidance in Complex Environment 346
Ying Wang, Huiting Zhang, Lu Yu, Kan Zhang, Xianghong Sun, and Thomas Plocher

Part IV: In-Vehicle Interaction

A Study and Evaluation on Route Guidance of a Car Navigation System Based on Augmented Reality 357
Kengo Akaho, Takashi Nakagawa, Yoshihisa Yamaguchi, Katsuya Kawai, Hirokazu Kato, and Shogo Nishida

Evaluation of Collision Avoidance Prototype Head-Up Display Interface for Older Drivers 367
Vassilis Charissis, Stylianos Papanastasiou, Lewis Mackenzie, and Sachi Arafat

The H-Metaphor as an Example for Cooperative Vehicle Driving 376
Daniel Damböck, Martin Kienle, Klaus Bengler, and Heiner Bubb

Factors for Representing In-Vehicle Roominess 386
Wonil Hwang, Nam-Hyo Kim, Hyeong-Joon Ahn, and Hee-Seok Jung

Analysis of Low-Floor Bus Passengers' Behavior Patterns Using Video Observation 391
Ji Yeon Kim, Hwan Hwangbo, Beom Suk Jin, Bong-Ha Hwang, Young Joo Moon, and Yong Gu Ji

The Effective IVIS Menu and Control Type of an Instrumental Gauge Cluster and Steering Wheel Remote Control with a Menu Traversal 401
Seong M. Kim, Jaekyu Park, Jaeho Choe, and Eui S. Jung

Assessing the Effect of a Power-Flow Gauge on Driving Behaviors Affecting Energy Consumption 411
Sang-Hwan Kim, Heramb Dandekar, Edgar Camez, and Heather Harrelson

In-Car Dictation and Driver's Distraction: A Case Study 418
Martin Labský, Tomáš Macek, Jan Kleindienst, Holger Quast, and Christophe Couvreur

Driver's Experience and Behavioral Patterns through the Observation of Commercial Vehicle Driving 426
Youngjae Lim, Sungjoon Park, Eui S. Jung, and Taeil Kim

Predicting the Effects of Time-Gaps for Adaptive Cruise Control (ACC) on Bus Driver Performance 435
Brian Tsang-Wei Lin and Sheue-Ling Hwang

Beginner Driver Support System for Merging into Left Main Lane...... 444
Yuki Nakamura and Yoshio Nakatani

Multimodal Interface for Driving-Workload Optimization 452
Hyesun Park, Jongwoo Choi, Hyeong-Joon Kwon, and Kyong-ho Kim

Part V: Social and Environmental Issues in HCI

Proposal of a Method for Promotion of Continuous Pro-Environmental Behavior with Easy Communication 465
Saizo Aoyagi, Tomoaki Okamura, Hirotake Ishii, and Hiroshi Shimoda

A Context Centric Approach to Utilize Social Media Services on Public Terminals 474
Micha Block, Jasmin Link, and Simon Thiel

Accessibility for Older Users through Adaptive Interfaces: Opportunities, Challenges and Achievements 483
Rob Edlin-White, Sue Cobb, Mirabelle D'Cruz, Anne Floyde, Sarah Lewthwaite, and Johann Riedel

Computer Usage and User Experience in Jordan: Development and Application of the Diamond Model of Territorial Factors 490
Fuad EL-Qirem and Gilbert Cockton

GooGreen: Towards Increasing the Environmental Awareness of Households 500
Ruud Mattheij, Lindsy Szilvasi, Lorraine de Beer, Kartini Rakiman, and Suleman Shahid

User Experience of Social Bookmarking Tools 510
Enric Mor, Nuria Ferran, Muriel Garreta-Domingo, and Juan-Antonio Mangas

Part VI: Emotions in HCI

ShoeBox: A Natural Way of Organizing Pictures According to User's Affinities 519
Bojan Blažica, Daniel Vladušič, and Dunja Mladenić

Toward Adapting Interactions by Considering User Emotions and Capabilities 525
Idoia Cearreta and Nestor Garay-Vitoria

A Haptic Emotional Model for Audio System Interface 535
Yuki Ichiyanagi, Eric W. Cooper, Victor V. Kryssanov, and Hitoshi Ogawa

Guess Who? An Interactive and Entertaining Game-Like Platform for Investigating Human Emotions 543
Muneeb Imtiaz Ahmad, Hassan Tariq, Mehreen Saeed, Suleman Shahid, and Emiel Krahmer

Adaptive Machine Learning Approach for Emotional Email Classification 552
K. Karthik and R. Ponnusamy

Designing Poetic Interaction in Space 559
Yi-Chu Lin, Huang-Ming Chang, and Rung-Huei Liang

Spectral Subtraction Based Emotion Recognition Using EEG 569
Jin-Hong Min, Hyeong-Oh Kwon, and Kwang-Seok Hong

Improving Human-Machine Interaction – A Non-Invasive Approach to Detect Emotions in Car Drivers ... 577
Michael Oehl, Felix W. Siebert, Tessa-Karina Tews, Rainer Höger, and Hans-Rüdiger Pfister

Emotion Recognition Using Biological Signal in Intelligent Space ... 586
Kanlaya Rattanyu and Makoto Mizukawa

Intentionality in Interacting with Companion Systems – An Empirical Approach ... 593
Dietmar Rösner, Rafael Friesen, Mirko Otto, Julia Lange, Matthias Haase, and Jörg Frommer

Multimodal Emotion Classification in Naturalistic User Behavior ... 603
Steffen Walter, Stefan Scherer, Martin Schels, Michael Glodek, David Hrabal, Miriam Schmidt, Ronald Böck, Kerstin Limbrecht, Harald C. Traue, and Friedhelm Schwenker

Author Index ... 613

Table of Contents – Part IV

Part I: HCI and Learning

A Web-Based Learning Environment to Support Chemistry 3
Candice Adams and Cheryl Seals

Introducing Mobility in Serious Games: Enhancing Situated and Collaborative Learning 12
Sébastien George and Audrey Serna

Visualization Framework for Computer System Learning 21
Eiichi Hayakawa, Yuuki Nakagawa, Hideharu Ochiai, Masahiko Fuji, and Yosuke Nishino

Associating Learners' Cognitive Style with Their Navigation Behaviors: A Data-Mining Approach 27
Yung-Chi Hsu and Sherry Y. Chen

The Design of Adaptive Error Feedback Music Ear-Training System with Image Cues 35
Yu Ting Hwang and Chi Nung Chu

Fuzzy Linguistic Modelling Cognitive / Learning Styles for Adaptation through Multi-level Granulation 39
Ilham N. Huseyinov

Method for Cultivating the "Inquiry-Mindset" Using the Information Access-Based Belief Bias Parameter 48
Kyoko Ito, Yuki Ito, and Shogo Nishida

Distance Education at the Graduate Level: A Viable Alternative? 58
Brian M. Jones, Andrea Everard, and Scott McCoy

Creating a New Context for Activity in Blended Learning Environments: Engaging the Twitchy Fingers 61
Jayne Klenner-Moore

Haptically Enhanced User Interface to Support Science Learning of Visually Impaired 68
Yueqing Li, Steve Johnson, and Chang Nam

Using Grounded Theory and Text Mining to Find Interesting Reading Materials for Slow EFL Learners 77
Yuh-Chang Lin, Chia-Ling Hsu, Mu-Hua Lin, Hsiao-Fang Yang, and Chao-Fu Hong

CAI Platform for Fundamental Geometric Training on Perspective Sketching 86
Ding-Bang Luh and Shao-Nung Chen

A Reading History Logger for Supporting Reading Habit Development 93
Yasuo Miyoshi and Takaaki Oobayashi

A Drawing Learning Support System with Auto-evaluating Function Based on the Drawing Process Model 97
Takashi Nagai, Mizue Kayama, and Kazunori Itoh

Interactions between Human and Computer Networks: EFL College Students Using Computer Learning Tools in Remedial English Classes 107
Ai-Ling Wang

Proposal of Collaborative Learning Support Method in Risk Communications 113
Hiroshi Yajima and Naohisa Tanabe

Evaluation of Online Handwritten Characters for Penmanship Learning Support System 121
Tatsuya Yamaguchi, Noriaki Muranaka, and Masataka Tokumaru

Facial Expression Recognition for Learning Status Analysis 131
Mau-Tsuen Yang, Yi-Ju Cheng, and Ya-Chun Shih

Part II: Health and Medicine Applications

An Enriched Understanding of Why the Environment and Individual Characteristics Are Important in Understanding Technology Utilization in Healthcare: An Evolutionary Psychology Perspective 141
Chon Abraham and Iris Junglas

A Real-Time Interactive MIDI Glove for Domicile Stroke Rehabilitation 151
Nizan Friedman, David Reinkensmeyer, and Mark Bachman

What Label Design of Ampule for Injection, Do You Want? 159
Hiroyuki Furukawa

The Design of an Interactive Stroke Rehabilitation Gaming System 167
Linda Harley, Scott Robertson, Maribeth Gandy, Simeon Harbert, and Douglas Britton

Therapeutic Category Improvement Method Based on the Words Appearing in Effect-Efficacy Description 174
Hirotsugu Ishida, Keita Nabeta, Masaomi Kimura, Michiko Ohkura, and Fumito Tsuchiya

Clinical Communication: Human-Computer and Human-Human Interactions 182
Saif Khairat and Yang Gong

Using Pen-Based Computing in Technology for Health 192
Hyungsin Kim, Young Suk Cho, and Ellen Yi-Luen Do

Using a Smart Phone for Information Rendering in Computer-Aided Surgery 202
Gaël Le Bellego, Marek Bucki, Ivan Bricault, and Jocelyne Troccaz

A Proposal of Contraindication Database for Medicines 210
Ryo Okuya, Hirotsugu Ishida, Keita Nabeta, Masaomi Kimura, Michiko Ohkura, and Fumito Tsuchiya

Results of the Usability and Acceptance Evaluation of a Cardiac Rehabilitation System 219
Cecilia Vera-Muñoz, María Teresa Arredondo, Ignacio Peinado, Manuel Ottaviano, José Manuel Páez, and Arturo Díaz de Barrionuevo

Construction and Analysis of Database on Outer Cases of Medicines 226
Hironori Yoshimi, Hiroki Muraoka, Akira Izumiya, Masaomi Kimura, Michiko Ohkura, and Fumito Tsuchiya

Part III: Business and Commerce

Are MIS Students Learning What They Need to Land a Job? 235
Andrea Everard, Brian M. Jones, and Scott McCoy

Promotion Project for Communication between Artisans and Consumers Supported by Media Technology 237
Ritsuko Izuhara, Sho Yokokawa, and Shinya Suzuki

Why Virtual Job Recruitment Is Not Well Accepted by Generation Y?—A Case Study on Second Life 245
Eleanor Loiacono, Soussan Djamasbi, Bengisu Tulu, and Oleg Pavlov

Investigating Online Advertising in Chile 255
Scott McCoy, Cristóbal Fernández Robin, and José Luis Cortés

Analysis of Customer Satisfaction on the Stiffness of Outside Panels of Passenger Cars 257
Ilsun Rhiu, Taebeum Ryu, Byungki Jin, and Myung Hwan Yun

Working toward Women's Economic Empowerment: Using Information and Communication Technology in Developing Areas to Market Traditional Crafts ... 266
Melissa Secore Levis

Socio Economic Psycho Knowledge Based Intelligent Agents for Automated e-Commerce Negotiation ... 274
P. Vijayaraghavan and R. Ponnusamy

Shopping Cart Interactive Program (SCIP) ... 285
Cyndi Wiley, Emmanuel Saka, Stefan Tauber, and Sunghyun R. Kang

Part IV: HCI in Complex Environments

An Analytical Alarm Flood Reduction to Reduce Operator's Workload ... 297
Jens Folmer, Dorothea Pantförder, and Birgit Vogel-Heuser

Self Replicating Robotic Strategies as a Catalyst for Autonomous Architectural Construction ... 307
Michael A. Fox

Development of Information Filtering Systems for Disaster Prevention ... 318
Yoshinori Hijikata, Tsutomu Yamanaka, Yuya Tanaka, and Shogo Nishida

Spatial Design, Designers and Users: Exploring the Meaning of Multi-party Service Cognition ... 328
Tom Hope, Mizuki Oka, Yasuhiro Hashimoto, and Myeong-Hee Lee

Toward an Understanding of a Computerized Monitoring System Failure: An Interpretive Approach ... 336
Nathan Johnson, Yibai Li, Fengchun Tang, and Saonee Sarker

Proposal of BCM Evaluation Method Based on Disaster Scenario Simulation ... 346
Ryuhei Kaneko and Yoshio Nakatani

Design of Communication Field for Leading to Satisfied Understanding: Example of High-Level Radioactive Waste Disposal in Japan ... 354
Hiroshi Kimura and Masashi Furukawa

Control Error Analysis of Computerized Operational Environment in Nuclear Power Plants ... 360
Seung Jun Lee, Jaewhan Kim, and Seung-Cheol Jang

uMeeting, an Efficient Co-located Meeting System on the Large-Scale Tabletop 368
Jie Liu and Yuanchun Shi

Enhanced User Experience in Managing Personal Finance 375
Cindy Lu

Experimental Investigation of Misuse and Disuse in Using Automation System 384
Akihiro Maehigashi, Kazuhisa Miwa, Hitoshi Terai, Kazuaki Kojima, Junya Morita, and Yugo Hayashi

Validating Video Analytics in Mission Critical Applications 394
Stephen J. Mitchell, Sukhpreet Gill, Steve Loveless, and Brent Auernheimer

Proposal of an Office Work Productivity Model Based on Short Pauses in Mental Tasks 403
Kazune Miyagi, Show Kawano, Hong Zhe Jin, Hiroshi Shimoda, and Hirotake Ishii

Restoration Support System for a Historic Textile Market Using Virtual Environment 413
Michiko Ohkura, Mizuki Konuma, Yuri Kogure, Sayaka Tanaka, Hitomi Ei, Akiko Sakai, Takashi Ishidou, and Yoko Watanabe

Part V: Design and Usability Case Studies

Investigating the Accessibility of Program Selection Menus of a Digital TV Interface 425
Pradipta Biswas and Patrick Langdon

Windows Positioning System: Aural Assistance Environment for the Aging in Windows Navigation 435
Chi Nung Chu

User Interactive Design for Digital TV Web Surfing 439
Chih-Fei Chuang, Nancy Huang, and Sheue-Ling Hwang

An End User and Environment Field Study for an Inclusive Design of Consumer Products 443
Thomas Fiddian, Chris Bowden, Mark Magennis, Antoinette Fennell, Joshue O' Connor, Pierre T. Kirisci, Yehya Mohamad, and Michael Lawo

Effects of Age Groups and Distortion Types on Text-Based CAPTCHA Tasks 453
Chih-Hsiang Hsu and Ying-Lien Lee

Evaluating Usability of Web-Based Electronic Government: Users' Perspective ... 456
Zhao Huang and Laurence Brooks

The Effects of Content Type and Presentation Style on User Experiences of Multimedia Content on a Tablet PC ... 466
Kari Kallinen, Jan Kallenbach, and Niklas Ravaja

Inherent Usability Problems in Interactive Voice Response Systems ... 476
Hee-Cheol Kim, Deyun Liu, and Ho-Won Kim

Effect of Aesthetic Design Elements on Tabletop Display Interaction ... 484
Hyunglae Lee, Hyunjin Shin, and Ji-Hyung Park

Effects of Presence on Causing Cybersickness in the Elderly within a 3D Virtual Store ... 490
Cheng-Li Liu and Shiaw-Tsyr Uang

A Development of Web-Based Player for Instructions Recorded with the Electronic Blackboard System IMPRESSION ... 500
Yuichi Ohkawa and Takashi Mitsuishi

Categorize Web Sites Based on Design Issues ... 510
Amin Rasooli, Fattaneh Taghiyareh, and Peter Forbrig

Interacting with Semantics and Time ... 520
Christian Stab, Kawa Nazemi, Matthias Breyer, Dirk Burkhardt, and Arjan Kuijper

Investigating Drag and Drop Techniques for Older People with Cognitive Impairment ... 530
Frédéric Vella, Nadine Vigouroux, and Pierre Rumeau

Part VI: Children and HCI

An Interface for Opportunistic Discovery of Information for Young People ... 541
Jamshid Beheshti and Andrew Large

Evaluating Leading Web Search Engines on Children's Queries ... 549
Dania Bilal and Rebekah Ellis

How Children Can Design the Future ... 559
Mona Leigh Guha, Allison Druin, and Jerry Alan Fails

Effects of Print-Storybooks and E-Storybooks with Reading Comprehension Strategies on Fifth Graders' Reading Comprehension Ability 570
Hsiu-Shuang Huang, Shang-Liang Chen, Yea-Mei Leou, Ho-Chuan Huang, Ching-Yu Yeh, Yun-Yao Chen, Chun-Lien Chen, and Ya-Ying Tseng

The Interaction of Children's Concepts about Agents and Their Ability to Use an Agent-Based Tutoring System 580
Alicia M. Hymel, Daniel T. Levin, Jonathan Barrett, Megan Saylor, and Gautam Biswas

A Comparison of Children's and Adults' Retrieval Performances and Affective Reactions When Using a Conventional Interface and an Information Visualization Interface 590
Andrew Large and Jamshid Beheshti

Following the Signs: Children's Use of Visual Cues to Facilitate Website Evaluation 599
Valerie Nesset

Development of Web-Based Voice Interface to Identify Child Users Based on Automatic Speech Recognition System 607
Ryuichi Nisimura, Shoko Miyamori, Lisa Kurihara, Hideki Kawahara, and Toshio Irino

Comparison of a 3-D Expression System and a Standardized IQ Test for Children 617
Akihiro Suzuki, Masayuki Wajima, Takashi Kawakami, and Tetsuo Okazaki

Exploring Children's Requirements for the Graphic Design of WebOPAC 627
Tengku Siti Meriam Tengku Wook and Siti Salwah Salim

Influence of Gender and Age on the Attitudes of Children towards Humanoid Robots 637
Fang-Wu Tung

Part VII: Playing Experience

Affective Videogames: The Problem of Wearability and Comfort 649
Andrea Bonarini, Fiammetta Costa, Maurizio Garbarino, Matteo Matteucci, Maximiliano Romero, and Simone Tognetti

Extraversion and Computer Game Play: Who Plays What Games? 659
Xiaowen Fang and Miaoqi Zhu

User Modeling Approaches towards Adaptation of Users' Roles to Improve Group Interaction in Collaborative 3D Games 668
Johanna Renny Octavia, Anastasiia Beznosyk, Karin Coninx, Peter Quax, and Kris Luyten

MusicTagger: Exploiting User Generated Game Data for Music Recommendation 678
Hannes Olivier, Marc Waselewsky, and Niels Pinkwart

The Influence of Social Experience in Online Games 688
Hua Qin, Pei-Luen Patrick Rau, and Song-feng Gao

Head-Pose Recognition for a Game System Based on Nose's Relative Position 694
Qingjie Zhao, Xiaoming Shi, and Yuxia Wang

Author Index 703

Part I

HCI International 2011 Keynote Speech

Technology-Mediated Social Participation: The Next 25 Years of HCI Challenges

Ben Shneiderman

Dept of Computer Science & Human-Computer Interaction Lab
University of Maryland, College Park, MD 20742 USA
ben@cs.umd.edu

Abstract. The dramatic success of social media such as Facebook, Twitter, YouTube, blogs, and traditional discussion groups empowers individuals to become active in local and global communities. Some enthusiasts believe that with modest redesign, these technologies can be harnessed to support national priorities such as healthcare/wellness, disaster response, community safety, energy sustainability, etc. However, accomplishing these ambitious goals will require long-term research to develop validated scientific theories and reliable, secure, and scalable technology strategies. The enduring questions of how to motivate participation, increase social trust, and promote collaboration remain grand challenges even as the technology rapidly evolves. This talk invites researchers across multiple disciplines to participate in redefining our discipline of Human-Computer Interaction (HCI) along more social lines to answer vital research questions while creating inspirational prototypes, conducting innovative evaluations, and developing robust technologies. By placing greater emphasis on social media, the HCI community could constructively influence these historic changes.

Keywords: social media, participation, motivation, social network analysis, user-generated content, Open Government, collective intelligence, collective action, community design, wikis, blogs, discussion groups, reader-to-leader framework.

1 Introduction

Generations of philosophers, psychologists, sociologists, and other social scientists have wrestled with deep questions about what motivates people to behave helpfully or harmfully. They also considered how to encourage people, organizations, and nations to resolve conflicts and how to inspire students, citizens, or employees to participate more creatively. Other important questions revolve around how trust grows, respect emerges, responsibility is accepted, and empathy is encouraged.

Contemporary researchers ask questions about what constitutes appropriate respect for privacy, when does online joking becomes bullying, or what constitutes fair use of others' creative work. These researchers also want to serve the growing numbers of community managers and social entrepreneurs who are seeking design guidelines, metrics of success, and strategies for stimulating social participation.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 3–14, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Getting validated guidance is gaining importance as social media enable much more frequent social contacts, contributions, and collaborations among larger numbers of people, covering a wider array of social situations. Advocates see this historical moment as a utopian opportunity that can restore failing economies, hasten business innovation, and accelerate citizen participation in government [1,2 3].

At the same time, critics complain that these technology-mediated relationships are necessarily shallower because they lack the richness of face-to-face contacts and are diffused across a larger number of people [4,5,6,7]. These critics worry about the deterioration of personal relationships, reduced capacity for groups to generate transformative social movements, higher likelihood of contentious public discourse, balkanized scientific communities, and much more. While business coordination, citizen participation, and international collaboration may be facilitated, there is a risk of reduced corporate loyalty, oppressive governments monitoring citizen communications, and destructive cyber-warfare.

2 Defining Technology-Mediated Social Participation (TMSP)

The increased use of social media has taken on new forms from personal blogs that allow lengthy thoughtful, often passionate, commentaries read by those with shared interests to the steady stream of 140-character tweets broadcast to hundreds of diverse followers and possibly retweeted to thousands more. Vigorous participation in social networks encourages awareness, responses, and sometimes serious discussions among "friends", while well-crafted YouTube videos can go viral when the web address is emailed, tweeted, posted to blogs, or mentioned on national television. The remarkable capacity of collaborative tools encourages massive coordinated efforts such as Wikipedia or beneficial collective intelligence projects such as film recommender systems, product review websites, or governmental knowledge sharing such as the U.S. State Department's Diplopedia. User-generated content sites also include photo sharing such as Flickr, music sharing, poetry, political essays, how-to, question-answering, open source software, and much more.

Controversies rage over the volume of benefits and the degree of harm produced by these user-generated content strategies. Utopian visionaries foresee empowered creative individuals, improved family communication, thriving resilient communities, outpouring of business innovation, citizen participation in policy decisions, and resolution of international conflicts.

Fearful critics complain about distracted attention, breakdown of family values, and digital Maoism that requires unpaid labor thus destroying jobs. They worry about increasingly contentious public discourse, uncontrolled spread of harmful cultural norms, and imposition of undesired economic policies on weaker international partners. Some critics disparage these social connections as lightweight and unable to produce committed participation that is necessary for social change [6]. They believe that social media only generate weak ties, while strong ties are needed to transformative activism. These claims may be true, but many respondents believe that large numbers of weak ties are a helpful and maybe necessary precursor to developing stronger ties among a narrower group of activists who ultimately produce potent changes.

Certainly the story of Jodie Williams is instructive and inspirational. She generated global email and discussion group communities to build strong ties that led to a successful movement to ban land mines, for which she received the Nobel Peace Prize. Another instructive example is the substantial social media response to the December 2009 Haitian earthquake. Not only did relief groups coordinate by way of social media, but millions of people texted a $10 contribution to the Red Cross in support of emergency responses. Never before had such a large amount of money been collected so quickly. Another form of social response was the rapid formation of Crisis Camp software marathons to generate software that helped map the disaster in areas where there had not been detailed street maps and software that produced translations from Haitian dialects for which there were no existing tools.

The benefits and harms from general social media are of broad interest and will produce extensive research from companies and e-commerce researchers, plus entertainment and social analysts who value these playful and discretionary technologies. At the same time there are important questions when these same social media are applied to national priorities and life-critical activities such as disaster response, community safety, health/wellness, energy sustainability, and environmental protection. These applications and contexts should become the focus of Technology-Mediated Social Participation (TMSP) research agendas.

An even larger circle of applications is included when the Open Government movement in many countries becomes part of the research agenda. In the United States, President Obama's Open Government Directive covering transparency, collaboration, and participation [8]. Data sharing at U.S. government web sites such as data.gov and recovery.gov have already changed agency practices and public discourse in many areas, but greater changes will come as participation is solicited through contests, challenges, volunteer invitations, expert review of proposed regulations, and national, state, and local service projects.

3 Setting a Research Agenda

Resolving the differing impressions about the benefits or harm of TMSP is more than an important challenge for community, business, and national leaders. It also leads human-computer interaction researchers to deal with profound scientific questions about individual behavior, collaborative strategies, community engagement, and international cooperation. By embracing these challenges, we can redefine HCI more broadly, maybe even signaling the change which a fresh name such as "human-community interaction" or even "human-social interaction (HSI)."

These shifts and name changes would refresh our research community with compelling research challenges that would lead us toward more profound questions. By embracing ambitious interface design goals and integrating new social science research questions, we would gain the chance to influence the still unfolding design of social media technologies and their applications. These opportunities bring trillion-dollar business possibilities and social transformation potentials that will shape civilization for centuries. Entrepreneurs will rush towards these possibilities, while scientific researchers from many disciplines will have a remarkable historic opportunity to develop profound theories about foundational aspects of individual, family, group, organizational, and national behaviors.

These questions have been studied for centuries, but the unique possibilities are that for the first time in history massive data about actual human activity is easily available for analysis. This information-abundant environment enables statistical analysis of billions of data points representing actual behavior, rather than a small sample survey of biased perceptions, constructed attitudes, or filtered reports of remembered activities.

The availability of abundant data is good news, but the algorithms for statistical, data mining, and machine learning methods, visualization tools, and methods for visual analytics are yet to be developed so as to study these compelling and deep science questions. Several research communities have identified this opportunity, most notably the promoters of web science [9]. Other sources include early efforts at social network analysis that go back almost a century, but only now are the software tools becoming available to do complex analysis and produce meaningful visualizations that show important patterns, clusters, relationships, anomalies, and outliers. A century of research is needed to develop novel mathematical methods and efficient algorithms to handle these new challenges, when data sets include millions and billions of vertices and edges.

Efficient algorithms are essential, but they may not be sufficient to give the rapid results that promote exploration. Just as specialized computer chips, known as Graphic Processing Units (GPUs), were needed to enable rapid manipulation of 3D environments represented by triangular meshes and enriched by visually compelling texture mapping, specialized computer chips, let's call them Social Processing Units (SPUs), may be necessary to handle network analysis computations. The algorithms for computing network metrics such as betweenness centrality, PageRank, or eigenvector centrality are difficult to implement on standard parallel computing architectures because most graphs defy clean data partitioning. Similarly, the growing ensemble of clustering algorithms that identify closely-connected communities require massive computational power as networks grow and linkages become denser. Finally, much greater computational power is needed to run the aggregation methods that simplify graphs so that they can be visualized by the growing family of layout strategies that users chose from to extract varying insights.

Early research by physicists and mathematicians produced attention-grabbing results describing scale-free and small-world networks generated by principles such as preferential attachment [10,11,12,13]. Then a new generation of computer scientists went further in integrating social science questions producing more applicable insights about strong vs. weak ties, sentiment prediction, and degree of participation [14,15,16]. As computational social scientists promote research in massive data sets [17], the community of researchers has grown dramatically, and funding is increasing from defense agencies, as well as national science agencies, national health funders, and an array of corporate sponsors.

There still is resistance to these new topics as entrenched disciplines skillfully lobby for support of their research agendas. They will not yield easily, so the HCI community and others who recognize this historic shift will have to collaborate effectively to compose a thoughtful and persuasive research agenda with realistic short-term and captivating long-term goals. Our request for massive funding shifts will be most effective if we engage with many disciplines, develop deep scientific theories and describe extreme technology challenges, all with well-documented linkages to societal benefits.

To help promote discussion of research agendas, we (myself, Jennifer Preece, and Peter Pirolli) obtained a U.S. National Science Foundation grant to conduct two 30-person workshops among leading academics, industry researchers, government staffers, and younger graduate students (www.tmsp.umd.edu). The first workshop was held in Palo Alto, CA on December 10-11, 2009 so as to attract West coast researchers and industry researchers and the second workshop was held in Arlington, VA on April 22-23, 2010 to facilitate participation by East coast academics, NSF staff and other government staffers.

Table 1. Potential domains of application of TMSP and the expected benefits. Adapted from [21].

Domain	Expected Benefit
Healthcare	Unite professionals and citizens in one center to gain information and to support and improve research and policy
Disaster response	Improve emergency response through citizen feedback and better response planning.
Energy	Facilitate creative thinking about energy alternatives and policies to bring new, environmentally friendly sources to the fore.
Education	Help make educational practices and policies more cost-effective.
Culture and diversity	Enhance understanding of the cultural variations both within and between societies.
Environment & climate	Enable broader understanding of the issues involved in the environment and climate change.
Citizen science	Promote the participation of citizens in areas of science where individuals can make useful contributions.
Economic health	Engage a broad base of citizens in thoughtful discussions about economic policies.
Globalization & development	Foster a better understanding of the emerging global economic and political realities.
Political participation	Increase informed political participation at all levels of government.
Local civic involvement	Cultivate increased understanding of ways to engage in local community activities.
Public safety	Encourage citizen participation in sharing information that can make their communities safer.

The workshops participants identified several themes and formed discussion groups spanning 6-9 months to develop six jointly authored articles, which were published in the November 2010 issue of *IEEE Computer* as the cover feature, with an introduction from the three workshop organizers [18]. The topics of the articles were scientific foundations [19], interface design issues [20], research infrastructure needs

[21], health/wellness possibilities [22], open government policies [23], and educational needs from K-12 through life-long learning [24].

In addition to health/wellness and open government applications, these detailed TMSP research plans highlighted applications in disaster response, energy, education, culture and diversity, environment & climate, citizen science, economic health, globalization & development, political participation, local civic involvement, and public safety (Table 1). The article on scientific foundations outlined new directions for theories, giving a sense of the scientific research opportunities that were neatly interwoven with practical problems.

4 New Theories and Innovative Research Methods

The range of TMSP theories needed is staggering, from *descriptive* theories that come from cleaned and aggregated data organized into meaningful insights to *explanatory* theories that present cause and effect patterns. These theories lay the foundations for *prescriptive* theories that provide guidelines and best practices for interface designers, community managers, and policy makers. In some situations *predictive* theories will be precise enough to forecast the evolution of social networks and the outcomes of collective action projects. Occasionally, deeper insights will lead to *generative* theories that suggest new social media strategies, novel methods for limiting malicious behavior, and new goals for collective action and international cooperation. These five flavors of theories are only a starting point, others forms of theories are likely to emerge as well to accommodate the breadth and depth of this vast research frontier.

Ideally, fresh theories lead to innovative research questions that require novel research methods [25]. The research methods for social media could bring the greatest change. The past 400 years of traditional science research in physics and chemistry has been governed by belief in the reductionist approach and replicable laboratory-controlled experimentation. Scientists would change a small number of independent variable, e.g. temperature or pressure, control other variables, e.g. electrical or magnetic fields, and measure the impact on dependent variables, e.g. resistance or expansion. The reductionist approach also influenced observational methods in botany, geology or astronomy, in which data was collected and analyzed to find relationships among variables that could be verified by independent researchers.

In the world of Technology-Mediated Social Participation, there may be new challenges for these traditional assumptions of reductionism and replicability. The variables of interest in TMSP include trust, empathy, responsibility, and privacy that are still hard to define and difficult to measure. Even frequently-discussed variables such as motivation, persuasion, self-efficacy, technology acceptance, and universal usability defy simple measurement beyond subjective scales that produce volatile and nuanced responses. Shifting to measurement of actual behaviors will be helpful in raising validity, but the tightly intertwined reactions of users means that context matters, familiarity is central, and results may be very different six months later when users are more or less trusting. The impact of privacy policy changes on trust for discretionary playful services such as film rating communities may be very different from life-critical systems such as disaster response or health discussion groups describing cancer treatment experiences.

If reductionism and replicability need to be redefined or replaced, how will journal reviewers revise their definitions of rigorous research? Can case studies move beyond hypothesis generation to become acceptable strategies to gather evidence that supports hypothesis testing? Maybe the examples of medical and business researchers could be useful, as they have developed standards for rigorous case study research that when repeated helps collect evidence in support of hypotheses [26]. Already, strategies such as Multi-dimensional Long-term In-depth Case studies (MILCs) are being applied to exploratory and discovery tools, such as in information visualization [27].

New forms of theories based on novel mathematics could emerge as they did in the early 20th century for quantum physics and statistical mechanics. Strange entanglements, maybe different from the quantum kind, are common in human experiences and statistical methods may nicely account for aggregated human behaviors, even as the actions of an individual are unpredictable.

The enduring questions of raising human motivation have taken on new importance in the age of social media. Wikipedia is a great success story because of its innovative strategies for motivating users to contribute their knowledge and to collaborate with others. But even in this success story, only one in a 1000 readers become registered contributors, and even fewer become regular collaborators who work together over weeks and months. Similarly, while there are billions of viewers of YouTube the numbers of contributors of content is small. Motivation or persuasion is an ancient human notion, but the capacity to study it on a global scale is just becoming a reality. The move from controlled laboratory experiments to interventions in working systems is happening because designers and researchers have enabled the capture of usage patterns on a scale never before possible.

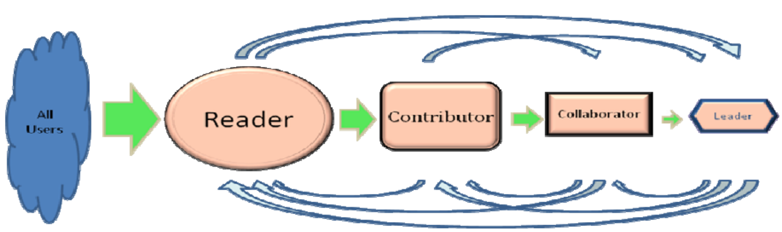


Fig. 1. The Reader-to-Leader Framework suggests that the typical path for social media participation moves from reading online content to making contributions, initially small edits, but growing into more substantive contributions. The user-generated content can be edits to a wiki, comments in a discussion group, ratings of movies, photos, music, animations, or videos. Collaborators work together over periods of weeks or months to make more substantial contributions, and leaders act to set policies, deal with problems, and mentor new users [28].

The Reader-to-Leader Framework [1] (Fig. 1) provides an orderly way of discussing motivational strategies and conducting research [28]. It suggests that users become readers by way of recommendations from friends, news media reports, web searches, and a hundred other techniques. Then readers may see contributions made by others and register so as to edit existing information or add new content, reviews, or ratings. The occasional contributor might evolve into a frequent contributor or go to the next step which is collaborating with others to plan new content. Some contributors become increasingly committed to the project and can develop strong relationships with other contributors. Then some contributors become engaged in governance, setting policy, dealing with problems, or mentoring newcomers. At each stage innovative entrepreneurs and researchers have developed interface design and motivational strategies such as showing the number of views of a video, enabling ratings of contributions, honoring richer collaborations, and empowering leaders.

Many other theories and frameworks are being proposed as commercial, government, and academic researchers rapidly expand their efforts. Traditional social science theories are being adapted to understand, predict, and guide designers who seek to increase trust, empathy, responsibility, and privacy in the online world [29, 30, 31]. Similarly, mathematical theories of network analysis are being enhanced to accommodate the distinctly human dynamics of online social systems. The shift from descriptive and explanatory theories that are based on statistical analyses and data mining to predictive and prescriptive theories that provide guidance for community managers is happening rapidly, but much work remains to be done.

5 Pursuing Reliable, Secure, and Scalable Technology

The past 40 years of computing technology have produced remarkable progress. Strong credit goes to the chip developers who made the rapid and sustained strides characterized by Moore's Law – doubling of chip density, speed, capacity every 18 months. Equal credit goes to the user interface designers who opened the doors to billions of users by creating direct manipulation interfaces based on carefully designed menus, effective graphical interfaces, convenient input devices, and comprehensible visual presentations.

The current agenda is rapidly moving to encompass the large-scale social media communities, such as the half billion users of Facebook and the four billion users of cell phones. Newer services such as Twitter have acquired more than 100 million users with billions of exchanges per month, but that is just the beginning. As individuals, organizations, companies, and governments increase their usage, the volume and pace of activity will grow bringing benefits to many users, but so will the impacts of service outages, privacy violations, and malicious attacks.

Developers now recognize the primacy of the user interface in determining outcomes, so there is increased research, training, and exploratory design. Simultaneously, there is a growth in tools for community managers to track, analyze, and intervene in social media networks to as to promote more positive outcomes. These tools will also be useful to researchers as they develop predictive models of network evolution, detect regions or growing or declining activity, and formulate theories of what motivates participation.

One such effort is the free, open source NodeXL Project (Network Overview for Discovery and Exploration in Excel), which is supported by Microsoft Research (www.codeplex.com/nodexl). This tool enables importing of social media networks from Outlook, Twitter, YouTube, Flickr, WWW, etc. into Excel 2007/2010, and then gives users powerful analysis tools, plus rich visualization support [32,33] (Fig. 2).

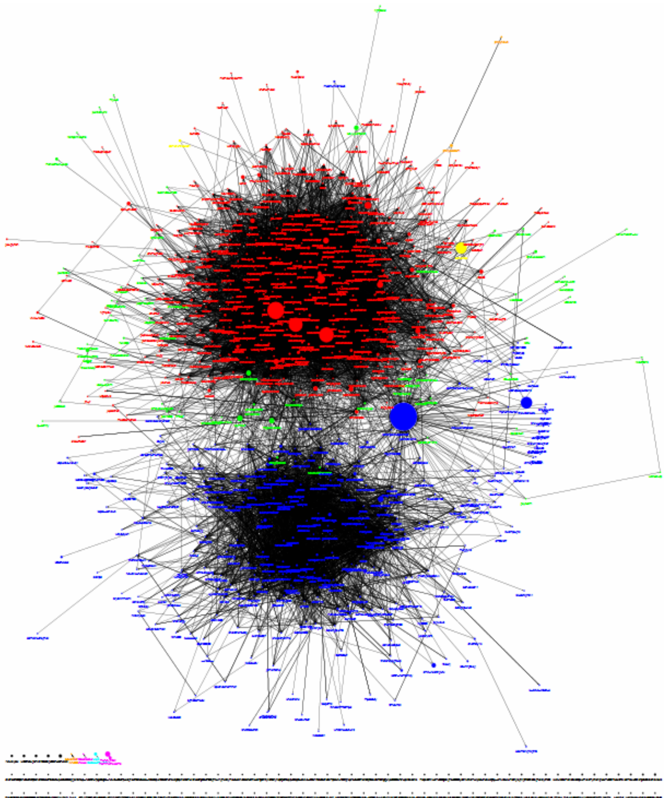


Fig. 2. This network shows connections among the Twitter users who mentioned #GOP when queried on January 8, 2011, with node size proportional to numbers of followers. The clusters are created by the patterns of connections (follows, replies, and mentions) among the authors in the graph. The clusters were based on Newman-Moore algorithmic analysis in which the Red cluster is composed of GOP supporters, while the Blue cluster contains critics and opponents of the GOP as indicated by the content of the tweets from each cluster. Other colored or shaped nodes are not strongly affiliated with either major cluster. (Created by Marc A. Smith using NodeXL http://www.codeplex.com/nodexl).

NodeXL was designed to speed learning by social-media savvy community managers and business professionals who already use Excel, as well as by undergraduate and graduate students who are learning social network analysis. By providing easy import of data from important social media tools, NodeXL dramatically expands the community of users who can carry out analyses that lead to actionable business insights and research studies. NodeXL provides a rich set of visualization controls to select color, size, opacity, and other attributes of vertices and edges. The variety of layout algorithms and dynamic query filters allows users to tune the display to their needs. Varied centrality metrics for directed and undirected graphs, as well as a growing number of clustering algorithms, support exploration and discovery. NodeXL is an ongoing project that will be supported through the emerging Social Media Research Foundation (www.smrfoundation.org).

6 Conclusion

These new TMSP research directions expand the scope of HCI research, shifting the emphasis from psychological to sociological issues, while engaging with new communities of algorithm developers, statistical data analysts who work with huge data sets, privacy protection researchers, and application specialists in new domains, especially e-government. TMSP also bring HCI researchers even more actively into the arena of mobile, ubiquitous, and pervasive technologies, while increasing the importance of attending to universal usability and international development. The possibilities for breakthrough science and transformational applications seem high, so there are a rapidly growing set of conferences and journals to serve these emerging topics [34]. While enthusiasm and optimism for TMSP is warranted, there should also be concern about the dangers of privacy violation, misleading information, malicious attacks, lack of universal usability, and failures during peak loads, such as during disasters. Since there are also dangers of use of this potent technology by criminals, terrorists, racial hate groups, and oppressive governments, our community will have to develop ethical standards and work to promote positive social values.

Acknowledgments. Thanks to the community of U.S. National Science Foundation workshop participants (www.tmsp.umd.edu) supported by grant IIS-0956571, the NSF Social Computational Systems grant IIS-0968521, the NodeXL team (www.codeplex.com/nodexl) supported by Microsoft External Research, and the University of Maryland Human-Computer Interaction Lab. I appreciate comments on the draft from Constantine Stephanidis, Gavriel Salvendy, and Jennifer Preece.

References

1. Tapscott, D., Williams, A.D.: MacroWikinomics: Rebooting Business and the World, Portfolio, New York (2010)
2. Shirky, C.: Cognitive Surplus: Creativity and Generosity in a Connected Age. Penguin Press, New York (2010)
3. Surowiecki, J.: The wisdom of the crowds. Random House, New York (2004)
4. Mozorov, E.: The Net Delusion: The Dark Side of Internet Freedom. Public Affairs, New York (2011)
5. Carr, N.: The Shallows: What the Internet is Doing to Our Brains. Norton, New York (2010)
6. Gladwell, M.: Small change: Why the revolution will not be tweeted, The New Yorker (October 4, 2010)
7. Lanier, J.: You Are Not a Gadget: A Manifesto. Knopf Publishers, New York (2010)
8. U.S. White House, Open Government Directive (December 9, 2009), http://www.whitehouse.gov/open/documents/open-government-directive

9. Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., Weitzner, D.: Web Science: An interdisciplinary approach to understanding the Web. Communications of ACM 51(7), 62–69 (2008)
10. Watts, D.J.: A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences of the United States of America 99(9), 5766–5771 (2002)
11. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 393(4), 440–442 (1998)
12. Newman, M., Barabasi, A.-L., Watts, D.J. (eds.): The structure and dynamics of networks. Princeton University Press, Princeton (2006)
13. Newman, M.E.J., Watts, D.J.: Scaling and percolation in the small-world network model. Physical Review E 60, 7332–7342 (1999)
14. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, New York (2010)
15. Leskovec, J., Huttenlocher, D., Kleinberg, J.: Predicting positive and negative Links in online social networks. In: International Conference on World Wide Web (WWW) (2010), ACM Digital Library
16. Kleinberg, J.: The convergence of social and technological networks. Commun. ACM 51(11), 66–72 (2008)
17. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., et al.: Computational Social Science. Science 323(5915), 721–723 (2009)
18. Pirolli, P., Preece, J., Shneiderman, B. (eds.): Technology-Mediated Social Participation (Cover Feature with 7 articles), IEEE Computer, vol. 43 (11), pp. 20–67 (November 2010)
19. Kraut, R., Maher, M.L., Olson, J., Malone, T.W., Pirolli, P., Thomas, J.C.: Scientific foundations: A case for Technology-Mediated Social Participation theory. IEEE Computer 43(11), 22–28 (2010)
20. Chi, E.H., Munson, S., Fischer, G., Vieweg, S., Parr, C.: Advancing the design of social participation systems. IEEE Computer 43(11), 29–35 (2010)
21. Olson, G., Mark, G., Churchill, E., Rotman, D.: Visions of what is possible with sharable socio-technical infrastructure. IEEE Computer 43(11), 37–43 (2010)
22. Hesse, B., Hansen, D., Finholt, T., Kellogg, W., Thomas, J.C.: Social participation in Health2.0. IEEE Computer 43(11), 45–52 (2010)
23. Bertot, J.C., Jaeger, P.T., Munson, S., Glaisyer, T.: Engaging the public in open government: Social media technology and policy for government transparency. IEEE Computer 43(11), 60–67 (2010)
24. Lampe, C., Resnick, P., Forte, A., Yardi, S., Rotman, D., Marshall, T., Lutters, W.: Educational priorities for Technology-Mediated Social Participation. IEEE Computer 43(11), 53–59 (2010)
25. Shneiderman, B.: Science 2.0. Science 319, 1349–1350 (2008), `http://www.sciencemag.org/cgi/content/full/319/5868/1349`
26. Yin, R.K.: Case Study Research: Design and Methods, 4th edn. Sage, Thousand Oaks (2008)
27. Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: Multi-dimensional In-depth Long-term Case Studies. In: Proc. Beyond time and errors: novel evaluation methods for Information Visualization, Workshop of the Advanced Visual Interfaces Conference, ACM Digital Library, pp. 1-7 (2006)
28. Preece, J., Shneiderman, B.: The Reader-to-Leader Framework: Motivating technology-mediated social participation. AIS Transactions on Human-Computer Interaction 1(1), 13–32 (2009), `http://aisel.aisnet.org/thci/vol1/iss1/5/`
29. Golbeck, J.: Weaving a Web of Trust. Science 321(5896), 1640–1641 (2008)

30. Preece, J.: Online Communities: Designing Usability, Supporting Sociability. John Wiley and Sons, Chichester (2000)
31. Maloney-Krichmar, D., Preece, J.: A multilevel analysis of sociability, usability and community dynamics in an online health community. ACM Transactions on Computer Human Interaction 12(2), 1–32 (2005)
32. Hansen, D., Shneiderman, B., Smith, M.A.: Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann Publishers, San Francisco (2011)
33. Smith, M., Shneiderman, B., Milic-Frayling, N., Mendes-Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A., Gleave, E.: Analyzing (social media) networks with NodeXL. In: Proc. Communities & Technologies Conference (2009)
34. Center for Advanced Study of Communities and Information, Publication Venues, http://casci.umd.edu/Publication_Venues (accessed on February 19, 2011)

Part II

HCI Design

Towards a Cognitive-Based User Interface Design Framework Development

Natrah Abdullah, Wan Adilah Wan Adnan, and Nor Laila Md Noor

Faculty of Computer and Mathematical Sciences
Universiti Teknologi Mara
Shah Alam, Malaysia
{natrah,adilah,norlaila}@tmsk.uitm.edu.my

Abstract. This paper discusses the theoretical framework underlying the studies of cognitive-based user interface design of heritage tourism website. Multiple Resource Theory under cognitive psychological study is used particularly in developing UI taxonomy of museum website. MRT highlights on three components which are perceptual modality, visual channel and code of processing. Components of MRT are applied extensively into user interface dimensions identification by emphasis on user interface support functions. As a result three components are propose; format, structure and representation. These components can be used to provide insights into area of HCI with taxonomy of UI for museum websites. Cognitive-based UI framework is proposed and presented with aims to assist in the design and development of the taxonomy.

Keywords: User Interface Design, Cognitive.

1 Introduction

User interface (UI) is a new medium to allow museum collections to be exhibited since online museum were being introduced. While Schweibenz [24] emphasis on effective way to display and easy to interact of museum artifact, result from exploratory study on museums website [25] indicates that there is lack of features on presenting and interacting with museum artifact on current museums websites. An absent of the features will effect on learning during exploring museum collections. Therefore, the lacking of capability of UI to educate or to persuade user make UI design in this domain needs to be explored.

The existence of study on user needs and their capability to process information help to design UI. In effect we must understand how people think and reason about cognitive concept before we make decisions about designing UI. The process of identifying of UI may be strongly supported if theoretical grounding is used as a platform [26]. Thus, this paper provides a theoretical study of UI dimensions in heritage tourism website. Besides suggests a foundation related to cognitive, this paper also highlights an important impact on UI design.

The paper is organized as follows. Section 2 discusses literature review to user interface design implemented in web application and we state a problem. Then, we describe the range of reasonable assumptions for the problem. Sections 3 and 4, a

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 17–24, 2011.
© Springer-Verlag Berlin Heidelberg 2011

potential UI dimensions for supporting UI design on web application is proposed, based on Multiple Resource Theory. Section 5 outlines the main potential benefits employing such a dimensions in a web application environment and discusses further research and development steps in this direction.

2 Literature Background

In this paper discussion on UI dimensions proceeds from a substantial research literature grounded in cognitive science. Cognitive science provides a unique perspective to the study of UI involving research from psychology and applied in the field of human computer interaction (HCI). This integration will provide new insights into the major factors influencing UI design and, consequently, could result in more efficient searching and browsing in museum website environments.

2.1 User Interface Design

Designing user interface can be done using techniques such as task-based design, model-based design, user-centered design and usage-centered design. Unlike task-based, model-based and usage-centered, which focus on system and process, the user-centered design focuses on user's characteristics in the design of an interface [1] and put an accent on making usable to the user [2]. Discussions on UI based on user characteristic emerged during the 1970s with the development of graphical user interface. Features associates with this type of interface led to major improvements as compared to command-line interface. The success of the innovation, from a cognitive perspective, was based on the finding that humans attempt to understand computers as analogical extensions of familiar activities [3]. This led to availability of desktop interface and direct manipulation interaction.

With internet, a revolution is occurring in the development of user interface. Whilst Welie [4] design interface with aim to allow user to focus on the task at hand and reduce the amount of overhead knowledge required communicating effectively with the website, Langdon and friends [15] emphasize on user capability and limitation during processing information. In addition, Chen et al [5] proposed a theoretical framework of UI underlying cognitive processes. Chen and his colleagues revealed that users' cognitive styles significantly influence their reaction to the user interface in terms of format, accessibility, and structure. These user interface designs are mostly discuss through literature study and there has been little discussion about theoretical studies in the area of UI designs. Thus, this study plan to explore the theoretical foundation for supporting UI design particularly looking at UI in cognitive perspectives.

2.2 Why Cognitive Study?

It has been observed that the goal of user interface design is to reduce the gap between users and system [6] and the cognitive-based design is a good way to achieve this goal [7]. Cognitive-based design can improve recall of information [8] and make applications more satisfying and easier to learn [9]. Callaway and his colleagues [10] report on good recall of information during museum visit based on cognitive aspect. They argue that cognitive aspect is important in designing museum exhibition. In addition [11] stress on human's cognitive characteristic play a key role in designing satisfy user interface. They concluded that understanding the components and the associated skills becomes a focus of attention and helps in designing usable UI design. Similar conclusion have drawn by Schneiderman [6] and he included cognitive aspect as one of the guidelines to deliver satisfy user interface. Thus, consideration on cognitive aspect is important in enhancing usable user interface.

Cognitive is an important characteristic that influences the effectiveness of user activity. Graziola and friends [12] have studied on users' reactions to interface design and found that cognitive significantly influence user reaction to user interface in terms of user experience. They concluded that any interactive application played on museum helps in enhancing experience to the users. However, interactivity may effect system performance and the design should consider user cognitive load. Thus, consideration on cognitive aspect is important in enhancing effective communication between user and website.

Research on human cognition can provide established conceptual frameworks toward investigation of the ability to acquire knowledge involved in the use of UI of Web environment. A theory which can support of cognitive aspect is Multiple Resource Theory. Multiple resource theory is used for predicting resource demand when multiple tasks are concurrently executed [13]. These predictions have significant impact in the area of user interface design because most practical applications are carried out in multi-task environments. The main theoretical foundations of MRT originally were established by Wickens' and Baddeley's work [13], which provided the basis and perspective needed to guide the work described in user interface design.

3 Multiple Resource Theory

According to Basil [13] the multiple resource theory (MRT) is proposed by Wickens. The theory helps to identify on UI components which aim to support cognitive aspect. User interface components are identified based on three different dimensions: perceptual modality, visual channels, and code of processing.Three components of MRT and related UI component are discuss in the following subsections.

3.1 Perceptual Modality

In MRT, perceptual modality is associated with the human channels of perception [18]. Perception involves human senses like textual content, visual content, and sound content components [14]. In inter-human communication, modalities are being used to express meaningful conversational contributions include speech, gestures, touch, facial expressions and smell [14]. Fikkert and friends [16] refer modality as defined by Bunt and Beun as an attribute that indicates the mode, manner, or form of something that the human sense employed to process information. Recently, McGAnn [17] discussed on perceptual modality and claimed that perceptual modalities is relate to "mode of presentation" of a particular attention. Thus, perceptual modality is related with form in which the information is presented. For our study, format dimension of UI is proposed to be used as dimension related to perceptual modality.

3.2 Visual Channel

A visual channel is nested dimension within visual resources, distinguishing between focal and ambient vision [13]. Focal vision supports object recognition in reading text and recognizing symbols and ambient vision is responsible for perception of orientation and movement [18]. The focal and ambient vision appears to be associated with different resource structures [13][18] in the sense being associated with different types of information processing. Visual channel is thus used to recognize the information representation. For our study, representation dimension of UI is proposed to be used as dimension related to perceptual modality.

3.3 Codes of Processing

In MRT, code of processing is related to how user process information [18]. Two different methods of processing are recommended by Huthmann [19]. The recommendation is based on spatial ability and categorical ability. While, spatial ability is the capability for processing and imaging movements or other changes in visual object, categorical ability is the capability for processing and imaging movements or changes in verbal information. According to Langdon [14] analog/spatial processing maintains visual materials such as pictures and diagrams, while categorical/symbolic processes sustains verbal information. For our study, structure dimension of UI is proposed to be used as dimension related to code of processing.

4 Research Framework

Based on MRT, cognitive-based research framework is proposed. This research framework consists of three dimensions of UI identify through an aiding concept. The important of each component relate to each of MRT dimension is illustrated in Figure 1.

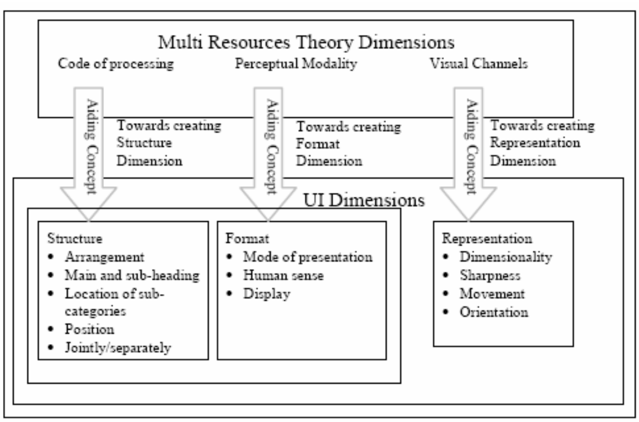


Fig. 1. Research Framework

4.1 Aiding Concept

Aiding concept is introduced by Greef and Neerincx [21]. Cognitive factor, support concept and support function are important in aiding concept. Aiding concept is used to derived UI dimensions. Users benefit from aiding, because the defined UI dimensions can invoke less demand cognitive processing. Table 2 shows the UI dimensions derived through the three levels of requirements.

Table 1. Proposed UI support Function Dimensions

MRT dimensions	Support concepts	UI support function dimensions
Visual Channels	Support the response selection and response execution	Representation • Sharpness • Dimensionality • Movement • Orientation
Perceptual Modalities	Support perception	Format • Mode of presentation • Human sense • Jointly/separately • Display
Code of processing	Support working memory	Structure • Arrangement • Main and sub-heading • Location of sub-categories • Position

4.3 Format Dimension

Format is a basic unit of the user interface space. In some cases these are sounds or textual and in other cases, images, either static or dynamic. Static type uses to display of electronic text. Dynamic type can serve a wide variety of iconic uses. Individual is tendency to view information either in static or dynamic format. However, to recognize objects, user must have cognitive ability for matching its features to image descriptions. Verbalisers tend to use verbal presentations to illustrate information when thinking, and prefer words rather than images. Imagers tend to use images to illustrate information when thinking, they understand visuals [20]. Verbalizes are superior at working with verbal information whereas imagers are better at working with visual and spatial information. Imager users view information in the form of photographs, diagrams, graphs, charts or basic animations. A verbalizer user tends to choose textual fashion such as written word or spoken.

4.4 Representation Dimension

Representation would be support response selection and response execution during information processing stage. During response selection, user would be attracted to choose object that directional to appropriate to in choose and easy to be exercised. Implementation on object happens when users choose to manipulate that object. Possible actions of manipulations a visual components include enlarge, zooming and rotation. Visual component permits object with characteristics that describe their structure and behavior [22]. Behavior will be indicated in a form of dimensionality. The dimensionality will determine whether object is 1D, 2D or 3D. Dimensionality will determine manipulation activity of user. Studies on this dimensionality have a comprehensive survey in HCI field and many advantages stated such as performance, experience and preferences.

4.5 Structure Dimension

Structure dimension will take place when user tries to understand a current situation. During processing stages, user will use their knowledge and experience and try to capture any knowledge kept in their mind. If the user have an experienced about the issue then it is easy to them to understand. Graff and [23] discussed on tendency for individuals to process information either as an integrated whole or in discrete parts of that whole. Analytical person process information from parts to the whole; able to establish meaningful structures, sequential, conceptually oriented, and prefer individualized learning. Wholists person process information from the whole to parts, they are factually oriented and affective, not highly organised and less likely to impose a meaningful organisation when there is no structure.

5 Conclusion

This paper explores the important of impact of cognitive aspects in UI design. MRT is explored and research framework is proposed using the MRT. In our proposed we highlight, to identify dimensions is important in museum for presenting object. There are several important implications of our research for research and practice. First, we used some existing concepts of cognitive to understand UI web environment. Second we integrated internal and external perspectives related to cognitive study to offer a holistic view of strategically UI development by using theoretical propositions suggested by MRT in forming the theoretical framework of UI dimensions.

Future work implies that this cognitive understanding of UI design may be further accommodated by the FI/FD dimension influences for the localization process of the UI features. In addition, our framework could be further empirically verified by researchers interested in this area of research. This may be done with the theoretical testing process research method by using practical heuristic UI design guideline relating to theoretical propositions highlighted in this study. Moreover, the theoretical building method reflected in this study that conducted by focusing on inductive reasoning may provide future empirical work toward validating the theoretical propositions for UI dimensions. Our research reflected in this paper is an effort to offer some theoretical understanding of UI dimensions by the adaptation of cognitive aspect. Furthermore, our work could be used as a starting point for conducting empirical studies to uncover the dynamics of UI dimensions.

References

1. Brown, S.: Let's Be Specific: Predicting Return On Investment in On-line Learning Activity Design. In: Trant, J., Bearman, D. (eds.) Proceedings of Museums and the Web 2007, Archives & museum Informatics, Toronto (2007)
2. Klett, F.: The Challenge in Learning Design Concepts: Personalization and Adaptation in Virtual Arrangements. In: ITHET 6th International Conference. IEEE, Los Alamitos (2005)
3. Yun, L., Yufen, C., Yingjie, W.: Cognition theory-based research on adaptive user interface for geo-visualization system. In: Proc. 12th Int. Conf. on Geoinformatics – Geospatial Information Research: Bridging the Pacific and Atlantic (2004)
4. van Welie, M.: Task-Based User Interface Design, thesis (2001)
5. Chen, S.Y., Magoulas, G.D., Macredie, R.D.: Cognitive styles and users' responses to structured information representation. International Journal Digital Library 4, 93–107 (2004)
6. Schneiderman, B., Plaisant, C.: Designing the User Interface: Strategies for Effective Human-Computer Interaction, 4th edn. Addison-Wesley, Reading (2004)
7. Fujita, H., Hakura, J., Kurematu, M.: Cognitive Evolutionary Techniques in User Interface Design Based on Human Aspect Modeling. International Journal of Computers 2(1) (2008)
8. Katrina Pariera, B.A.: Information Literacy On The Web: How Visual And Textual Cues Contribute To Website Credibility Assessments. Thesis (2009)
9. Oviatt, S.: Human-Centered Design Meets Cognitive Load Theory: Designing Interfaces that Help People Think. ACM, New York (2006)
10. Callaway, C., Kuflik, T., Not, E., Novello, A., Stock, O., Zancanaro, M.: Personal Reporting of a Museum Visit as an Entrypoint to Future Cultural Experience. In: IUI 2005 and ACM (2005)
11. Fikkert, W., D'Ambros, M., Bierz, T., Jankun-Kelly, T.J.: Interacting with Visualizations. In: Kerren, A., Ebert, A., Meyer, J. (eds.) GI-Dagstuhl Research Seminar 2007. LNCS, vol. 4417, pp. 77–162. Springer, Heidelberg (2007)
12. Goren Bar, D., Graziola, I., Pianesi, F., Zancanaro, M.: Dimensions of Adaptivity in Mobile Systems: Personality and People' s Attitudes. In: Proceedings of Intelligent User Interfaces IUI 2005, San Diego, pp. 223–230, January 200531–62 (2006)
13. Basil, M.D.: Multiple Resource Theory II: Application To Television Viewing. Communication Research 21, 208–231 (1994)
14. Langdon, P., Persad, U., John Clarkson, P.: Developing A Model of Cognitive Interaction for Analytical Inclusive Design Evaluation. Interacting with Computers 22, 510–529 (2010)
15. Jacobson, S.: Learning Style and Lawyering: Using Learning Theory to Organize Thinking and Writing. Journal of the Association of Legal Writing Directors 2, 27–73 (2004)
16. Fikkert, W., Bierz, T., D'Ambros, M., Jankun-Kelly, T.J.: Interacting with Visualizations. In: Kerren, A., Ebert, A., Meyer, J. (eds.) GI-Dagstuhl Research Seminar 2007. LNCS, vol. 4417, pp. 77–162. Springer, Heidelberg (2007)
17. McGann, M.: Perceptual Modalities: Modes of Presentation or Modes of Interaction? Journal of Consciousness Studies 17(1-2), 72–94 (2010)
18. Wickens, C.D.: Multiple Resources and Mental Workload. Human Factors 50(3), 449–455 (2008)
19. Huthmann, A.: Presence-Dependent Performance Differences Between Virtual Simulations And Miniature Worlds, thesis (2009)

20. Riding, R., Rayner, S.G.: Cognitive styles and learning strategies. David Fulton, London (1998)
21. Neerincx, M.A., de Greef, H.P.: Cognitive Support: Extending Human Knowledge and Processing Capacities. Human-Computer Interaction 13, 73–106 (1998)
22. Nguyen, G.T., Rieu, D.: Multiple Object Representations. In: Proc. 20th ACM Computer Science Conference, Kansas City (Mo) (1992)
23. Davies, J., Graff, M.G.: The wholist-analytic cognitive style: A matter of reflection. Personality and Individual Differences 41(6), 989–997 (2006)
24. Schweibenz, W.: The Virtual Museum: New perspectives for museums to present objects and information using the Internet as a knowledge base and communication system. (1998), http://www.informationswissenschaft.org/download/isi1998/14_isi-98-dvschweibenz-saarbruecken.pdf (retrieved September 6, 2009)
25. Natrah, A., Wan Adilah, W.A., Nor Laila, M.N.: A Framework of Web-based Cultural Tourism Information System Evaluation. In: The 2009 International Joint Conferences on e-CASE and e-Technology, pp. 655–665 (2009)
26. Wan Abdul Rahim, W.M.I., Nor Laila, M.N., Shafie, M.: Incorporating the Cultural Dimensions into the Theoretical Framework of Website Information Architecture. In: Aykin, N. (ed.) HCII 2007. LNCS, vol. 4559, pp. 212–221. Springer, Heidelberg (2007)

A Design Science Framework for Designing and Assessing User Experience

Sisira Adikari, Craig McDonald, and John Campbell

Faculty of Information Sciences and Engineering,
University of Canberra, ACT 2601, Australia
{Sisira.Adikari,Craig.McDonald,John.Campbell}@canberra.edu.au

Abstract. User Experience is a well recognized factor in design and evaluation of artifacts in Human-Computer Interaction. There are many user experience models reported in the literature to reflect this status. Techniques and instruments for managing user experience are still not sufficient. In this paper, we discuss design science research and important user experience models reported in the literature and propose an integrated design science framework for designing and assessing user experience. We also present the results of an experimental study to validate our proposed framework and the instrument employed.

Keywords: Design Science, User Experience.

1 Introduction

User Experience (UX) aims at gaining a more comprehensive understanding of users interactive experiences with products or systems around new concepts like pleasure [1], fun [2], aesthetics [3] and hedonic qualities [4]. Norman [5] describes the UX as encompassing all aspects of users interaction with a product. The distinction between usability and user experience has been recognised, for example, even a product with good usability can generate negative use experiences causing dissatisfaction while a product with bad usability can generate positive experiences or satisfaction [6], and while bad usability can break a good product, good usability is insufficient to create a good experience [7]. Accordingly, the good usability does not guarantee good UX, and it depends on the total interactive experience of a user with a product or system in meeting user needs and expectations. Although, there is no universal definition of user experience [7], the concept of UX has been widely adapted in many areas of design without it being clearly defined or deeply understood [8,9]. However, there are many views of user experience from the literature. According to Forlizzi and Battarbee, user experience is associated with a wide variety of meanings without cohesive theory of experience for the design community [10]. Hassenzahl and Tractinsky see user experience as ranging from traditional usability to beauty, hedonic, affective or experiential aspects of technology use [11].

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 25–34, 2011.
© Springer-Verlag Berlin Heidelberg 2011

A newer ISO standard [12] defines the UX as a persons perceptions and responses that result from the use or anticipated use of a product, system or service emphasizing two main aspects: use and anticipated use. This definition matches the description of experience given by McCarthy and Wright [13]. Moreover, the results of a recent research study show that ISO definition of user experience is in line with the view by most respondents about the subjectivity of UX [9]. ISO standard also points out that user experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after use and Usability criteria can be used to assess aspects of user experience [12]. UX assessment is an evaluation of the users interactive experience of a product, system or service. Accordingly, use (i.e., actual interaction experience), anticipated use (i.e., pre-interaction experience such as needs and expectations), and after use (post-interaction experience) is equally important for consideration in UX assessments. We consider that total UX consists of these three views and it is shown as a model in Figure 1.

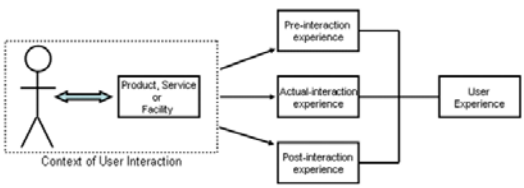


Fig. 1. Total user experience model

In this paper, we briefly discuss important user experience models reported in the literature and propose an enhanced usability model integrated with design science research framework for user experience design and assessment. We also present the results of an experimental study to validate our proposed model and the instrument employed.

2 Design Science Research

Design Science (DS) is a research paradigm [14] , [15] , [16] which aims at creating and evaluating innovative artifacts that address important and relevant organizational problems. The design science involves the purpose driven creation of artifacts and the introduction of these artifacts into a natural setting [15]. The main focus of design science research is to create and evaluate new and improved IT artifacts as a solution for relevant organisational problems and to generate new knowledge to the body of the scientific evidence. Hevner has presented a detailed process of performing design science research by means of three design science research cycles [17]. Figure 2 show the existence of three design science research cycles in information systems research framework reported in [18]. According to [16], these three research cycles must be present and clearly identifiable in any design science research project.

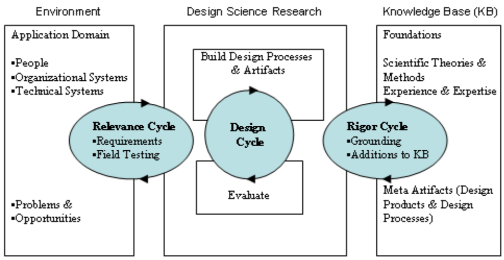


Fig. 2. Design science research cycles

The relevance cycle captures the problem to be addressed and connects the contextual Environment with Design Science Research activities. Thus, the relevance cycle provides the requirements for the research (problems and opportunities) as inputs. It also defines the acceptance criteria for design solutions and artifacts which feed back as outputs from Design Science Research to the Environment for implementation. The suitability of artifacts in Environment is assessed by field testing to determine whether additional iteration of the relevance cycle are needed.

As shown in Figure 2, the Knowledge Base consists of foundations, scientific theories and methods, experience and expertise and meta-artifacts. The rigor cycle connects the Design Science Research activities with the Knowledge Base to receive foundations for rigorous design science research to ensure its innovation. Moreover, the rigor cycle feedbacks new knowledge to the Knowledge Base to update and enrich it. The internal design cycle is central to the Design Science Research where the hard part of the actual design science research is carried out. Inputs to the Design Science Research are: requirements from the relevance cycle and the design and evaluation theories and methods which are drawn from the rigor cycle. The design cycle iterates between the core activities of building and evaluating the design artifacts to generate design alternatives and evaluating the alternatives rigorously and thoroughly against requirements until a satisfactory design is achieved. Finally, design artifacts and research contributions are output into the relevance cycle for implementation in Environment and field evaluations, and to updating the Knowledge Base respectively.

Design Science and Human-Computer Interaction share two important common grounds: design and evaluation. User experience design and evaluation is concerned with building HCI artifacts for contextual requirements and evaluating those against pre-specified assessment criteria. In practical terms, HCI artifact creation and evaluation follow an iterative approach similar to that explained in Design Science Research. Accordingly, we propose that the DSR framework shown in Figure 2 is a well suited framework that can be applied for designing for UX and UX assessment. Our proposed framework is shown in Figure 3 where the user interaction with a system in context is emphasized with 'UX Theories', 'Experimental Designs' and 'Evaluation Artifacts' included in the Knowledge Base.

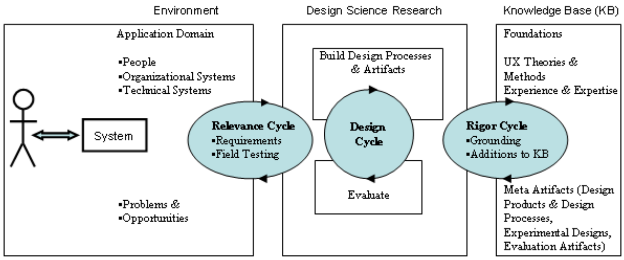


Fig. 3. Proposed design science research framework for designing for user experience and user experience assessment

3 User Experience Models

There are many user experience models reported in the literature explaining the aspects of user experience. Many of these models recognize the significance of usability on user experience. Specifically, these models seek to complement a purely functional analysis of user interaction with an account of the sensual, emotional, social and cultural aspects of peoples' relationships with technology [19].

In the context of interaction design, Preece, Rogers and Sharp [20] pointed out the difference between user experience goals and usability goals emphasizing that UX goals are more concerned with how users experience an interactive system from user perspective rather than assessing how useful or productive a system is from product perspective.Their model consists of six usability goals namely: efficient to use, effective to use, safe to use, having good utility, easy to learn, and easy to remember. They defined user experience goals as: satisfying, enjoyable, fun, entertaining, helpful, motivating, aesthetically pleasing, supportive of creativity, rewarding and emotionally fulfilling. According to them, usability goals are fundamental to the user experience and occurs as a result of achieving usability goals during an interaction. Importantly, the model does not include 'satisfaction' as a usability goal; instead, it considers as a UX goal. Taking this view further, Sharp, Rogers and Preece [21, p.26] described an enhanced version of UX goals that included additional positive as well as negative goals, namely: engaging, pleasurable, exciting, cognitively stimulating, provocation, surprising, challenging, enhancing sociability, boring, frustrating, annoying and cutesy. They described many of these goals as subjective qualities concerned with how a system feels to a user. They also highlighted that not all usability and UX goals will be relevant to the design and evaluation of an interactive product (or system) as some combinations will not be compatible.

According to Hassenzahl [22], the origins of UX can be seen from two different views: pragmatic quality and hedonic quality. Hassenzahl argues that the fulfilment of be-goals is the driver of experience, and that pragmatic quality facilitates the potential fulfilment of be-goals. Moreover, Hassenzahl [22] emphasizes that UX is an evaluative feeling of the user while interacting with a product with a shift of the attention from the product to feelings such as subjective side of product use. Another model [11] sees UX as a combination of three perspectives: beyond the instrumental, emotion and affect, and experiential and considers as a consequence of the user's internal state, the characteristics of the designed system and the context within which the interaction occurs. Having discussed models and theories of experience, Forlizzi and Battarbee [10] state that experience in interactive systems can be examined and modeled from three perspectives, namely: product-centered, user-centered, and interaction-centered. The definition of UX given by Hassenzahl and Tractinsky [11] also includes user perspective, product perspective and context of interaction perspective. Accordingly, UX is a mixture of these three perspectives: product-centred, user-centred and interaction-centred.

In a recent publication, Bevan [23] highlighted that how product attributes relate different aspects of UX. According to Bevan, direct measurement of actual experience of usage is difficult; the measurable consequences are the user performance, satisfaction with achieving pragmatic and hedonic goals, comfort and pleasure. User performance and satisfaction is determined by quality characteristics such as attractiveness, functional suitability, ease of use, learnability, accessibility and safety. Table 1 presents Bevan's view on how the measures of usability and UX are dependent on the product attributes that support different aspects of UX [23]. Moreover, Bevan points out that the measures of UX consequences can be represented by means of satisfaction in use, with a specific focus on pleasure, likeability and trust. According to Table 1, 'Pleasure' relates to Aesthetics, 'Likability and Comfort' relates to Appropriate Functions, Good UI Design, Learnability and Technical Accessibility and 'Trust' relates to Safe and secure Design.

Petrie and BevanBevan [24] have acknowledged that users of new technologies are not necessarily seeking just to complete a useful task, but also to amuse and entertain themselves. Accordingly, Petrie and Bevan consider that UX, as a concept, emerged to cover the components of users' interactions with, and reactions to, electronic systems that go beyond usability. Table 1 also shows that

Table 1. Factors contributing to UX consequences

Quality characteristic	UX	Functionality	User interface usability	Learnability	Accessibility	Safety
Product attributes	Aesthetic attributes	Appropriate functions	Good UI design (easy to use)	Learnability attributes	Technical accessibility	Safe and secure design
UX pragmatic do goals	To be effective and efficient					
UX hedonic be goals	Stimulation, identification and evocation					
UX: actual experience	Visceral	Experience of interaction				
Usability (= performance in use measures)	Effectiveness and Productivity in use: effective task completion and efficient use of time			Learnability in use: effective and efficient to learn	Accessibility in use: effective and efficient with disabilities	Safety in use: occurrence of unintended consequences
Measures of UX consequences	Satisfaction in use: satisfaction with achieving pragmatic and hedonic goals					
	Pleasure	Likability and Comfort				Trust

UX and usability are not two distinct concepts, but have interrelated aspects that contribute equally to providing the overall UX and usability of a system.

4 Extending Usability for User Experience Assessment

There are many definitions of usability by different authors such as [25,26,27] etc., where satisfaction is a commonly shared viewpoint. In their study, Lindgaard and Dudek [28] highlight that UX consists in some sense of satisfaction, and emphasize that aesthetics, emotion, expectation, likeability and usability all influence the interactive experience. In a recent research study reported [34] that user satisfaction received the highest agreement as a criterion of UX. Therefore, an enhanced usability model with more focus on satisfaction can be used to probe the satisfaction usability attribute in more depth to explored in identifying how the user feels about the interaction with a product, system or service. For UX assessments, we propose a usability model consisting of eight usability attributes: satisfaction, functional correctness, efficiency, error tolerance, memorability, flexibility, learnability and safety. This is an enhancement of the previous model presented [29], with an additional inclusion of the safety usability attribute. Our proposed usability model is shown in Figure 4 and it serves as an extension to Bevan's view on UX shown in Table 1. In our proposed usability model, we have not considered Technical Accessibility (TA) as a usability attribute and intend to study TA in more detail to determine its impact on usability and UX.

5 User Experience Experiment

As discussed, the satisfaction attribute can be used beyond aesthetics and conventional satisfaction to provide a broad view of experience on satisfaction in use in

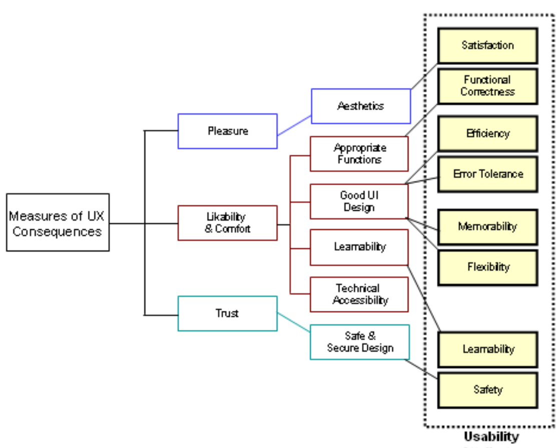


Fig. 4. Proposed usability model

achieving pragmatic and hedonic goals. This section details how we derived assessment criteria for the satisfaction usability attribute. There are many definitions for the term aesthetics, such as visual appeal [28], beauty in appearance [30], etc. A common viewpoint of these definitions is that aesthetics is related to pleasure and harmony, which human beings are capable of experiencing [30]. Accordingly, we consider Visual appeal and Pleasure in interaction as our first two assessment criteria. ISO view on UX shows that Expectations contribute to the UX; hence we chose Meeting expectations as our third assessment criterion. According to Ketola and Roto [31] , Frustration is a measure that could be used to assess UX obstacles in an interaction; therefore, Less frustration has been selected as an assessment criterion. Confusion or lack of understanding is considered as impacting on creating an overall positive user experience [33], hence we have included Less confusing terminology as an assessment criterion. Lastly, the assessment criterion Overall experience of using the system was included to receive a subjective opinion from the user on the overall experience of interaction.

First column of Table 2 shows the complete list of assessment criteria for each usability attributes in our proposed usability model shown in Figure 4. Particularly, the 'Satisfaction' usability attribute examines six UX aspects: visual appeal, pleasure in interaction, meeting expectations, less frustration, less confusing terminology, and overall experience of using the system.

5.1 Experimental Study

An evaluation questionnaire consists of 27 questions was developed based on the assessment criteria shown in first column of Table I. The experimental study was carried out in two separate phases. In the first phase, we choose a library information system as the Reference System and conducted a user experience assessment using the evaluation questionnaire with a group of 16 subjects (group G1). We deployed the design science research framework (see Figure 3) as the basis for the second phase of the project. At the beginning, an overview of the design science research framework was given to a team of two developers, and one user experience designer emphasizing the intent of the second phase of the experimental study, and the importance of the relevance and rigor aspects of the design science research framework. The first author acted as an observer to make sure that the team would carry out tasks as intended. For system design, the functional requirements of the Reference System were given to the team along with user experience assessment criteria. The team was then asked to develop new IT artifacts based on functional requirements and user experience assessment criteria through three research cycle approach specified in the design science research framework. Finally, the resulting system (Developed System) was subjected to a user experience evaluation with a different group of 16 subjects (group G2) using the same evaluation questionnaire.

5.2 Results

For both systems, user responses for the evaluation questionnaire were collected from both groups of subjects (G1 and G2). These data were statistically

Table 2. Results

UX Assessment Criterion	Reference System		Developed System		Significance (2-tailed)
	Total Mean	Std. Deviation	Total Mean	Std. Deviation	
Visual appeal	1.8750	1.0247	4	1.09545	0.000
Pleasure in interaction	2.1875	1.10868	3.3125	.60208	0.001
Meeting expectations	2.0625	.99791	3.6875	.94648	0.000
Less frustration	1.7500	.77460	3.6250	.88506	0.000
Less confusing terminology	2.2500	1.0000	3.2500	.93095	0.018
Overall experience of using the system	2.1250	.95743	3.1875	.83417	0.008
Completing tasks correctly	2.0000	1.03280	3.2500	.85635	0.001
Available facilities to meet user needs	2.4375	.96393	3.5000	.81650	0.003
Available information to make decisions	1.9375	.99791	3.5625	1.03078	0.000
Completing tasks quickly	2.0000	.73030	3.4375	.81394	0.000
Achieving expected outcome	1.8750	.71880	3.5625	.89209	0.000
Completing tasks easily	2.0625	.99791	3.3750	.80623	0.001
Causes fewer errors	2.0000	1.03280	3.8125	.91059	0.000
Clear error messaging for invalid conditions	1.8125	1.04682	3.4375	.81394	0.000
Error messages that inform which actions to take	1.9375	.92871	3.3750	.80623	0.000
Easiness to remember task steps	2.0000	.73030	2.9375	.85391	0.004
Needing to memorise task steps	2.4375	.89209	3.3750	.61914	0.004
Needing to access Help documents	2.4375	.81394	3.5	.51640	0.001
Alternative ways to perform tasks	2.1875	.91059	3.1875	.75000	0.002
Navigating back/forward between task steps	2.1875	1.04682	3.3750	.50000	0.002
User ability to cancel an operation	2.1250	.88506	3.5625	.62915	0.000
Ease of learning system operation	1.8125	.98107	3.5625	.96393	0.000
Clarity of system status	2.0000	.73030	3.8125	.83417	0.000
Knowing to do next during navig.	1.8125	.7500	3.9375	.92871	0.000
Fewer keystrokes	1.8125	.7500	3.9375	.92871	0.000
Sec. measures to protect personal info.	1.8125	.7500	3.9375	.92871	0.000
Sec. measures to protect user transactions	1.8125	.7500	3.9375	.92871	0.000

analyzed by means of independent samples T test to measure the significance of user perception for each assessment criterion over two systems. The aim was to determine if there was a significant difference of agreement of user responses for each assessment criterion over reference system and developed system. Table 2 shows the mean response values and respective standard deviations for each criterion over two systems and the resulting 2-tailed significance measure for each criterion.

According to Table 2, for each criterion except criterion 5, there is a significant difference of agreement from users for the Developed System with an associated higher level of mean value (>3). Even for the criterion 5, the mean value of developed system is > 3 (where as the mean value of the Reference System <3). This clearly shows that developed system has been rated as significantly better than reference system in terms of 27 UX assessment criteria.

6 Conclusions

In this paper, we have presented a number of approaches to UX modeling and derived a well-grounded assessment criteria that can be used for design for UX as well as conduct an effective UX assessment in HCI. We also presented an integrated approach based on design science research framework for design and assess UX. The main objective of the experimental study was to validate the integrated design science research framework for UX design and assessment.

The results of the experimental study clearly demonstrate that the integrated design science research framework was successful in generating positive results that highlighted the system differences.

References

1. Jordan, P.: The Grid: Designing Pleasurable Products: An Introduction to New Human Factors. Taylor & Francis, Abington (2000)
2. Blythe, M., Overbeeke, K., Monk, A., Wright, P.: Funology: from usability to enjoyment. Kluwer Academic Publishers, Dordrecht (2003)
3. Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. Interacting with Computers 13, 127–145 (2000)
4. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Human-Computer Interaction 19, 319–349 (2004)
5. Norman, D.: The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the answer. The MIT Press, Cambridge (1998)
6. Jokela, T.: When good things happen to bad products: where are the benefits of usability in the consumer appliance market? Interactions 11, 28–35 (2004)
7. Kuniavsky, M.: Smart Things: Ubiquitous Computing User Experience Design. Morgan Kaufmann, San Francisco (2010)
8. Law, E., Roto, V., Vermeeren, A., Kort, J., Hassenzahl, M.: Towards a shared definition of user experience. In: CHI, pp. 2395–2398 (2008)
9. Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: CHI, Boston, pp. 719–728 (2009)
10. Forlizzi, J., Battarbee, K.: Understanding experience in interactive systems. In: DIS 2004, Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 261–268 (2004)
11. Hassenzahl, M., Tractinsky, N.: User experience-a research agenda. Behaviour and Information Technology 25, 91–97 (2006)
12. ISO FDIS 9241-210: Human-centred design process for interactive systems. ISO (2009)
13. McCarthy, J., Wright, P.: Technology as experience. Interactions 11, 42–43 (2004)
14. Iivari, J.: A paradigmatic analysis of information systems as a design science. Scandinavian Journal of Information Systems 19, 39–63 (2007)
15. Baskerville, R.: What design science is not. European Journal of Information Systems 17, 441–443 (2008)
16. Hevner, A., Chatterjee, S.: Design science research in information systems - Theory and Practice. Springer, Heidelberg (2010)
17. Hevner, A.: The Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems 19, 87–92 (2007)
18. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Quarterly 28, 75–105 (2004)
19. Blythe, M.A., Wright, P.C.: Pastiche scenarios: Fiction as a resource for user centred design. Interacting with Computers 18, 1139–1164 (2006)
20. Preece, J., Rogers, Y., Sharp, H.: Interaction design: beyond human-computer interaction. Wiley, Chichester (2002)

21. Sharp, H., Rogers, Y., Preece, J.: Interaction design: beyond human-computer interaction. Wiley, Chichester (2007)
22. Hassenzahl, M.: User experience (UX): towards an experiential perspective on product quality. In: Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, pp. 11–15 (2008)
23. Bevan, N.: Classifying and selecting UX and usability measures. In: International Workshop on Meaningful Measures: Valid Useful User Experience Measurement, pp. 13–18 (2008)
24. Petrie, H., Bevan, N.: The evaluation of accessibility, usability and user experience. In: Stephanidis, C. (ed.) The Universal Access Handbook, pp. 20.1–20.14. CRC Press, Boca Raton (2009)
25. Nielsen, N.: The usability engineering. Academic Press, San Diego (1993)
26. Shneiderman, B.: Designing the user interface: strategies for effective human-computer interaction. Addison Wesley Longman, Boston (1998)
27. Dix, A., Finlay, J., Abowd, G.D.: Human-computer interaction. Prentice-Hall, Upper Saddle River (2004)
28. Lindgaard, G., Dudek, C.: What is this evasive beast we call user satisfaction. Interacting with Computers 15, 429–452 (2003)
29. Adikari, S., McDonald, C., Lynch, N.: Design science-oriented usability modelling for software requirements. In: 12th International Conference on Human-Computer Interaction: Interaction Design and Usability, pp. 373–382 (2007)
30. Lindgaard, L.: Aesthetics, visual appeal, usability, and user satisfaction: What do the user's eyes tell the user's brain 5, 429–452 (2007)
31. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human-Computer Studies 60, 269–298 (2004)
32. Ketola, P., Roto, V.: Exploring user experience measurement needs. In: 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement, VUUM (2008)
33. Tullis, T., Albert, W.: Measuring The User Experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann, San Francisco (2008)
34. Naumann, A.B., Wechsung, I., Schleicher, R.: Measurements and concepts of usability and user experience: Differences between industry and academia. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 618–626. Springer, Heidelberg (2009)

Objective and Subjective Measures of Visual Aesthetics of Website Interface Design: The Two Sides of the Coin

Ahamed Altaboli[1,2] and Yingzi Lin[1]

[1] Mechanical and Industrial Engineering Department,
Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA
[2] Industrial and Manufacturing Systems, Engineering Department,
University of Garyounis, Benghazi, Libya
Altaboli@gmail.com, yilin@coe.neu.edu

Abstract. The main purpose of this study is to compare objective layout-based measures of visual aesthetics with subjective questionnaire-based measures. Correlation analysis was used to carry out the comparison. Values for the tested objective measures were calculated for forty-two web pages already used in a previous study, for which subjective questionnaire scores (classical/expressive and VisA WI) were already available. Results showed significant correlations between many of the tested objective screen layout-based measures and subjective questionnaire-based measures related to order and layout of the screen. These findings suggest that the objective layout-based measures tested in this study can be used for overall assessments of visual aesthetics of websites and particularly for assessing aesthetic aspects related to the classical and the simplicity dimensions of website aesthetics.

Keywords: Aesthetics measures, Measure of Website aesthetics, Classical/expressive aesthetics, VisA WI, Visual aesthetics.

1 Introduction

Since its establishment in earlier 1980s, the field of the Human Computer Interaction (HCI) was mainly concerned with functionality and usability aspects of interactive computerized systems. However, recently, there was a new wave in the field emphasizing the importance of aesthetic aspects in HCI and interface design [13, 14, & 21].

1.1 Visual Aesthetics in Interface Design

The attention to the importance of aesthetics in interface design and its effect on users' impressions of usability of the system began with findings of Kurosu and Kashimura study [9]. Using different designs of an automated teller machine interface, they found high correlation between users' prior perception of usability (they called it apparent usability) and users' perception of visual aesthetics of the interface. Participants perceived the visually appealing interface designs as easier to use. Later, Tracinsky [29] repeated their experiment in a different context using more rigors approach, same high correlation was also found in all the tested cases. Furthermore,

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 35–44, 2011.
© Springer-Verlag Berlin Heidelberg 2011

this strong relationship between user perception of interface aesthetics and perceived usability remains intact even after actual use of the system [27]. Lindgaard et al. [12] showed that first impressions of perception of visual appeal of websites formed very quickly within 50 milliseconds. It remains stable even after a considerably longer exposure [27]. Phillips and Chapparro [23] found that users' impression of usability of websites is most influenced by the visual appeal of the site. Users rated sites with high visual appeal and low usability as easier to use, and gave lower rates to sites with low visual appeal and high usability.

Besides positive effect of aesthetics on perceived usability, some even argue that visually appealing interfaces might also have positive effects on performance. For example, Moshagen et al., [17] found significant effect of highly aesthetic websites on completion time in a low usability condition when participants completed search tasks. Sonderegger & Sauer [26] showed that visual appearance of cell phones had a positive effect on performance, leading to reduced completion time and number of errors for the visually appealing design.

With recognition of the effect of visual aesthetics of interface and computer screen design on users' overall evaluation of the system, the next natural step in the design process is to develop tools to assess and measure visual aesthetics.

1.2 Measures of Interface Aesthetics

In general, two approaches to measure interface aesthetics can be distinguished in the literature. The first is an objective approach relating screen design features and layout elements to the users' perception of visual aesthetics (e.g. [3 & 20]). The second one is a subjective approach, utilizing questionnaire-based instruments to measure users' perception of visual aesthetics (e.g. [11].

Objective Screen Layout-based Measures. This approach represents a bottom-up procedure. It has its roots in the rationalistic philosophical view of aesthetics [22]. This approach comprises the concept of "beauty in the observed object"; i.e. human perception of beauty is based on the order and organization of the various components constructing the object. It is concerned with determining what features in the interface design triggers users' precipitin of aesthetics of the interface. It also tries to explore the possibility of expressing changes in such features using numerical values and use these numerical values to assess users' perception of interface aesthetics. Methods in this approach are motivated by earlier aesthetic measures developed by Birkhoff [5], Tullis' quantitative techniques for evaluating screen design [32], and Gestlest theory for visual design [6 & 19].

Supports of this approach [3 & 19] argue that developing such measures can be very helpful in many design situations. They can be extremely helpful in early stages of design. They can assist in preparing design alternatives and can reduce the number of prototypes that will undergo tests with human users in later stages of design. However, they are not meant to be replacements to human designers, but are intended to serve as numerical tools to help designers and researches evaluate different design alternatives without the need to use human participants. Moreover, they can provide researchers with quantitative tools that can help in systematically study different design aspects and give a numerical basis for direct comparing of different design proposals. These measures can also be useful in cases where on-the-fly designs are needed for non professional designers as in online tools for designing websites [10].

Methods in this approach can be divided into two categories; one that simply uses numerical counts of visual features on the screen (like: number of objects, number of images …etc) and relates them to users' perception of aesthetics. The second one uses mathematical formulas to express more sophisticated visual design features and concepts (like: symmetry, balance …etc) and relate them to users' perception of aesthetics.

Simple counts measures. Visual features used in this categories includes number of constructing elements or blocks and chunks of information on the screen [3, 4, & 15], number of images [3, 4, 7, & 15], image size and font size [7 & 25], JPEG file size of screenshots of websites [31]

Formularized measures. Methods in this category argues that physical layout of visual objects on the screen may play a role in users' perception of aesthetics. The procedure involves expressing visual design features (like symmetry, balance, unity …etc) using mathematical formulas and combine calculated values for all features to build an overall measure that would reflect aesthetic level of the interface design.

One of such measures is the model developed by Ngo et al. [19]. The model consists of fourteen proposed measures of screen aesthetics: balance, symmetry, equilibrium, unity, sequence, density, proportions, cohesion, simplicity, regularity, economy, homogeneity, rhythm, and order. The value of each measure can be calculated using formulas based on the layout of visual objects on the screen. The average of all these measures represents the overall aesthetic value of the screen.

In a practical application of the model, Zain et al. [33] designed a computer application to incorporate five measures (balance, equilibrium, symmetry, sequence, and rhythm) of the fourteen measures. They used the software to evaluate language learning web pages.

Bauerly and Liu [3 & 4] developed mathematical formulas to measure and test effects of symmetry and balance on interface aesthetics. They developed their formulas based on a microscopic view that compare screen elements in question pixel by pixel, as opposed to the macroscopic view used by Ngo et al. [20] that compare higher level elements such as specific objects or shapes. However, similar results were obtained in both studies, which suggest that there are no practical differences between the two sets of formulas. Lai et al. [10] utilized the quantitative measures of symmetry and balance used by Bauerly and Liu [3 & 4] to quantitatively analyze the aesthetics of a text-overlaid image such that a best position for overlaying the texts on a background image can be obtained automatically. The two measures were evaluated against participants' subjective rating of visual aesthetic appeal in cases of color and monochrome images.

Subjective Questionnaire-based Measures. Supporters of this approach claim that the complexity and interrelated relationships among the screen design elements make it difficult to use them to quantitatively measure aesthetics [11]. It would be more convenient to use questionnaire-based instruments to measure users' subjective perception of aesthetics. Two widely accepted of such instruments are: the classical and expressive instrument developed by Lavie and Tractinsky [11], and the Visual

Aesthetics of Website Inventory (VisAWI) tool developed by Moshagen and Thielsch [16]. Both were designed to measure perceived visual aesthetics of websites.

Lavie and Tractinsky [11] found two dimensions of the perceived website aesthetics, termed “classical aesthetics” and “expressive aesthetics”. The classical aesthetics dimension emphasizes orderly and clear design and is closely related to many of the usability and interface design rules and guidelines. The expressive aesthetics dimension is linked to the designers’ creativity and originality and to the ability to break design conventions. These two dimensions were the basis for developing quantitative questionnaire-based instrument to measure website interface aesthetics. The classical dimension includes the items “aesthetic”, “pleasant”, “symmetric”, “clear”, and “clean”, while the expressive aesthetics includes the items “creative”, “fascinating”, “original”, “sophisticated”, and “uses special effects”.

VisAWI was constructed to serve as a new tool to measure perceived website aesthetics. It was designed to provide a tool that would cover border aspects of perceived websites aesthetics that weren't adequately presented in early instruments. The instrument is based on four interrelated facets of perceived visual aesthetics of websites: simplicity, diversity, colorfulness, and craftsmanship. Simplicity comprises visual aesthetics aspects such as balance, unity, and clarity. It is closely related to the classical aesthetics dimension. The Diversity facet comprises visual complexity, dynamics, novelty, and creativity. It is closely related to the expressive aesthetics dimension. The colorfulness facet represents aesthetic impressions perceived from the selection, placement, and combination of colors. Craftsmanship comprises the skillful and coherent integration of all relevant design dimensions. Each of the first two facets is presented by five items in the questionnaire, while each of the last two facets has four items. For full list of items of both questionnaires see [11 & 16].

1.3 Purpose of the Study

The main concern of this study is to investigate the possibility of finding significant correlations between objective and subjective measures of website aesthetics. The hypothesis is that significant correlations could be -at least- found between the objective layout-based measures and items related to screen layout in the subjective questionnaire-based measures (classical measure in the classical/expressive questionnaire and simplicity measure in the VisA WI questionnaire).

Forty-two web pages already used in a previous study [16] to develop and validate the VisA WI questionnaire and compare it with classical and expressive aesthetics questionnaire will be used in this study. Values of selected objective measures will be calculated for these 42 web pages and compared to subjective questionnaire scores (VisA WI and classical/expressive) already available in [16]. Correlation analysis will be used to carry out the comparisons.

The reason why these 42 web pages are utilized in this study is that they cover a wide variety of websites with different levels of visual aesthetics. In addition, questionnaire scores (for both VisA WI and classical/expressive tools) for a relatively large sample size are already available for these pages.

2 Results and Discussion

2.1 Selected Objective Measures for the Study

Formularized. Eight of the measures suggested by Ngo et al. [18] were selected for this study. The reasons for selecting these measures are that they are widely accepted and have already being used and tested in many previous studies [1, 18, 24, & 33]. The eight selected measures are: symmetry, balance, unity, sequence, simplicity, density, economy, and rhythm. The formulas developed by Ngo et al. [18] will be used to calculate values for the selected measures (for full details of how to calculate these measures and their definitions see [18, 19, & 20]). Table 1 shows descriptive statistics for the calculated values using the formalized selected measures for the 42 web pages. Possible range of values of each measure is zero to one, where zero represents the lowest aesthetic level and one represents the highest level.

Table 1. Descriptive statistics for the selected formularized measures for the 42 web pages

Measure	Min	Max	Average	Standard deviation
Symmetry	0.761	0.985	0.873	0.046
Balance	0.516	0.950	0.792	0.105
Unity	0.157	0.684	0.393	0.145
Sequence	0.750	1.000	0.970	0.082
Simplicity	0.077	0.273	0.152	0.044
Density	0.091	0.977	0.453	0.230
Economy	0.050	0.250	0.105	0.040
Rhythm	0.257	0.792	0.627	0.109

Counts. Five measures were selected in this category, namely: number of visual objects on the screen, number of different sizes of visual objects, number of images, number of different font types used in the web page, and JEPG file size of screenshot of the webpage. Number of objects, number of images, and JEPG files size have already been tested in previous studies [2, 3, 16, 14 & 24]; all with results indicting some sort of a relationship between these measures and users' perception of visual aesthetics. Number of different sizes of visual objects is one of the input parameters in Ngo et al. formulas for unity, economy, and simplicity; authors were interested in testing it as a single feature. Number of different font types has been selected based on earlier observation. Descriptive statistics for the calculated values for the selected count-based measures for the 42 web pages are given in Table 2.

The calculated values for the above selected measures for all the 42 web pages will be compared with scores of the classical/expressive and VisA WI questionnaires (see [16] for questionnaire scores given to the 42 web pages).

Table 2. Descriptive statistics for the selected count-based measures for the 42 web pages

Measure	Min	Max	Average	Standard deviation
No of objects	6	22	11.9	3.9
No of different sizes of objects	4	20	10.7	3.6
JEPG file size*	50	251	170.8	44.4
No of different font types	1	6	2.8	1.3
No of images	0	12	4.3	3.1

* in Kbytes

2.2 Correlation Analysis

Correlation analysis was used to investigate possible relationship between the selected objective and subjective measures of visual aesthetics of websites. Values of all the selected objective measures were calculated for the 42 web pages, scores of subjective questionnaire-based measures were obtained from [16]

Table 3 shows correlation coefficients between the formalized measures and questionnaire scores for the 42 web pages. From the table, one can see that all significant correlations are with questionnaire items related to screen layout. The objective measures of unity, simplicity, and economy are significantly correlated with the classical and the simplicity measures; both containing items related to visual layout and clarity of the design. No significant correlations were found between the formalized measures and questionnaire scores related to the expressive aesthetics.

An unexpected result is the lack of significant correlations between symmetry and balance, and any of the questionnaire scores. This contradicts with results of earlier studies showing significant effects of symmetry and balance on visual interface aesthetics [1, 3, 4, 10, 18, & 30]. Possible explanation could be found by looking at findings of an earlier study conducted by authors of the current study [2]. The study investigated effects of the measures of balance, unity, and sequence and their interactions on perceived visual aesthetics. Findings showed that with each pair of measures the effect of one measure is larger at high values of the other measure; with the low values the effect is very small. Looking back at Table 3, it can be seen that values of symmetry, balance, and sequence are much higher than values of measures with significant correlations (unity, simplicity, and economy). Symmetry, balance, and sequence have values larger than 0.761, 0.516, and 0.750 respectively with averages of 0.873, 0.792, and 0.970 respectively. On the other hand, measures with significant correlations have smaller values; Unity with values less than 0.684 and an average of 0.393, simplicity and economy with values less than 0.300 and averages of 0.152 and 0.105 respectively.

Since symmetry, balance, and sequence have higher values than unity, simplicity and economy, therefore, the later three will have higher effects on perceived aesthetics. This might explain why only these three measures have significant correlations. Nevertheless, the other case of lower values of symmetry, balance, and sequence should also be investigated to confirm this explanation. Also, Can the high levels of symmetry and balance witnessed in the current study be considered as a typical characteristic of all websites designs? Or is it just a coincidence with the 42 web pages used in the study?

Another point worth mentioning is that only measures that have number of different sizes of objects as an input parameter in their formulas (unity, simplicity, and economy) produced significant correlations. The measures of density and rhythm that do not have this parameter in their formulas didn't produce significant correlations, though have low values.

Table 4 shows correlation coefficients between the simple counts-based measures and questionnaire scores for the 42 web pages. Significant correlations were found between number of objects and number of different sizes with both the classical and the simplicity measures. This wasn't surprising, since these two features (no of objects and no of different sizes) are the main input parameters in the formulas used to calculate values of the formalized-based measures that showed significant correlations (unity, simplicity, and economy; for full details of the used formulas see [18, 19, & 20]). These significant correlations point out to clear negative effects of increasing number of objects and number of different sizes on perceived visual aesthetics of websites. The questions now are: Is there an optimal value or a threshold for these two features? How much can both of them be lowered? Do they affect each other?

No strong correlations were found between JEPG file size, number of different font types, and number of images with any of the classical and the simplicity measures. However, an interesting result is the noticeable high and significant correlations found between number of different fonts and the expressive and the diversity measures. Does this indicate that number of different fonts may affect visual aesthetics? More examinations are needed to clarify this.

3 Conclusions

The main goal of this study was to test the possibility of that significant correlations could be found between the objective layout-based measures of website visual aesthetics and the subjective questionnaire-based measures related to order and organization of visual objects on the screen. Values for the selected objective measure were calculated for forty-two web pages already used in a previous study [16] for which subjective questionnaire scores (classical/expressive and VisA WI) were already available. Correlation coefficients between the objective measures and questionnaire scores were calculated. Results showed significant correlations between many of the selected objective screen layout-based measures and the subjective questionnaire-based measures related to order and layout of the screen.

This suggests that the objective layout-based measures tested in this study could be used to generally assess the overall visual aesthetics of websites and particularly aesthetic aspects related to classical and simplicity dimensions of website aesthetics.

However, several issues still need to be considered when interpreting findings of this study. First, the formulas used to calculate values of the objective layout-based measures don't include effect of colors, although, Ngo et al [17] suggested adding effects of colors as part of the balance measure, but, it is still not clear how to express

Table 3 . Correlations between objective formularized -based measures and subjective questionnaire based measures

Measure	Classical/expressive			VisA WI				
	Classic	Expressive	Total	Simplicity	Diversity	Colorfulness	Craftsmanship	Total
Symmetry	-0.162	-0.185	-0.216	-0.065	-0.265	-0.360	-0.411	-0.302
Balance	0.064	0.064	0.080	0.136	-0.001	0.100	-0.111	0.044
Unity	0.449*	0.128	0.385**	0.569*	0.107	0.152	0.359*	0.363*
Sequence	0.279	0.062	0.229	0.313	0.131	0.297	0.167	0.269
Simplicity	0.370**	0.154	0.344**	0.461*	0.201	0.082	0.386**	0.344**
Density	0.275	0.112	0.255	0.243	0.215	0.055	0.311	0.247
Economy	0.560*	0.248	0.529*	0.653*	0.322**	0.141	0.428*	0.475*
Rhythm	-0.177	0.004	-0.121	-0.188	0.016	-0.036	0.036	-0.061

* Significant at 0.01, ** significant at 0.05

Table 4 . Correlations between objective simple count-based measures and subjective questionnaire based measures

Measure	Classical/expressive			VisA WI				
	Classic	Expressive	Total	Simplicity	Diversity	Colorfulness	Craftsmanship	Total
No of objects	-0.500*	-0.201	-0.461*	-0.583*	-0.227	-0.131	-0.408*	-0.413*
No of different sizes of objects	-0.630*	-0.284	-0.598*	-0.722*	-0.363**	-0.181	-0.505*	-0.542*
JEPG file size	-0.338**	0.023	-0.223**	-0.333	-0.011	0.038	-0.123	-0.142
No of different font types	-0.333**	0.600*	0.103	-0.257	0.399*	-0.172	-0.047	-0.019
No of images	-0.251	0.195	-0.066	-0.224	0.203	0.082	-0.016	0.002

* Significant at 0.01, ** significant at 0.05

effects of colors using numerical values. Second, for many of the relationships suggested by the significant correlations found in this study to be fully confirmed, further testing using more rigorous procedures is needed. Finally, since web pages were used in this study, findings of this study are only applicable to visual aesthetics of websites. Nonetheless, it would be interesting to see how these findings would work with other types of interfaces and screens (e.g. cell phones).

Acknowledgements. The authors would like to thank Morten Moshagen and Meinald T.Thielsch (authors of reference [16]) for providing the screenshots and questionnaire scores for the 42 web pages. This research has been supported by NSF through a CAREER program (Award # 0954579).

References

1. Altaboli, A., Lin, Y.: Experimental Investigation of Effects of Balance, Unity, and Sequence on Interface and Screen Design Aesthetics. In: Blashki, K. (ed.) Proceedings of The IADIS International Conference Interface and Human Computer Interaction 2010, Freiburg, Germany, pp. 243–250. IADIS Press (2010)
2. Altaboli, A., Lin, Y.: Investigating the Effects of Screen Layout Elements on Interface and Screen Design Aesthetics. Technical Report (unpublished). Intelligent Human-Machine Systems Lab, Mechanical and Industrial Engineering Department, Northeastern University, Boston, MA, USA (2010)
3. Bauerly, M., Liu, Y.: Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. Int. J. Human-Computer Studies 64, 670–682 (2006)
4. Bauerly, M., Liu, Y.: Effects of Symmetry and Number of Compositional Elements on Interface and Design Aesthetics. International Journal of Human- Computer Interaction 24(3), 275–287 (2008)
5. Birkhoff, G.: Aesthetic Measure. Harvard University Press, Cambridge (1933)
6. Chand, D., Dooley, L., Tuovinen, E.: Gestalt Theory in Visual Screen Design – A New Look at an Old Subject. Presented at the Seventh World Conference on Computer in Education, Copenhagen, Denmark. Australian Computer Society, Inc. (2002)
7. Djamasbia, S., Siegelb, M., Tullisb, T.: Generation Y, web design, and eye tracking. Int. J. Human-Computer Studies 68, 307–323 (2010)
8. Hartmann, J., Sutcliffe, A., De Angeli, A.: Towards a theory of user judgment of aesthetics and user interface quality. ACM Transactions on Computer-Human Interaction 15(4), 1–30 (2008)
9. Kurosu, M. and Kashimura, K.: Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: CHI 1995: Conference Companion on Human Factors in Computing Systems, Denver, Colorado, United States, pp. 292–293 (1995)
10. Lai, C., Chen, P., Shih, S., Liu, Y., Hong, J.: Computational models and experimental investigations of effects of balance and symmetry on the aesthetics of text-overlaid images. Int. J. Human-Computer Studies 68, 41–56 (2010)
11. Laviea, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. Int. J. Human-Computer Studies 60, 269–298 (2004)

12. Lindgaard, G., Fernandez, G., Dudek, C., Brown, J.: Attention web designers: You have 50 milliseconds to make a good impression. Behavior & Information Technology 25(2), 115–126 (2006)
13. Liu, Y.: Engineering aesthetics and aesthetic ergonomics: theoretical foundations and a dual-process research methodology. Ergonomics 46, 1273–1292 (2003)
14. Liu, Y.: The aesthetic and the ethic dimensions of human factors and design. Ergonomics 46, 1293–1305 (2003)
15. Michailidou, E., Harper, S., Bechhofer, S.: Visual Complexity and Aesthetic Perception of Web pages. In: SIGDOC 2008, Lisbon, Portugal, September 22-24 (2008)
16. Moshagen, M., Thielsch, M.T.: Facets of visual aesthetics. International Journal of Human-Computer Studies 68(10) (2010)
17. Moshagen, M., Musch, J., Göritz, A.S.: A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. Ergonomics 52, 1311–1320 (2009)
18. Ngo, D., Byrne, J.: Application of an aesthetic evaluation model to data entry screens. Computers in Human Behavior 17, 149–185 (2001)
19. Ngo, D.C.L., Teo, L.S., Byrne, J.G.: Evaluating Interface Esthetics. Knowledge and Information Systems (4), 46–79 (2002)
20. Ngo, D.C.L., Teo, L.S., Byrne, J.G.: Modelling interface aesthetics. Information Sciences 152(1), 25–46 (2003)
21. Norman, D.: Emotional design: Why we love (or hate) everyday things. Basic Books, New York (2004)
22. Reich, Y.: A model of aesthetic judgment in design. Artificial Intelligence in Engineering 8(2), 141–153 (1993)
23. Phillips, C., Chapparro, C.: Visual Appeal vs. Usability: Which One Influences User Perceptions of a Website More? Usability News 11(2) (2009)
24. Salimun, C., Purchase, H.C., Simmons, D.R., Brewster, S.: Preference ranking of screen layout principles. In: The 24th BCSHCI 2010, Abertay Dundee, September 6-10 (2010)
25. Schmidt, K.E., Bauerly, M., Liu, Y., Sridharan, S.: Web Page Aesthetics and Performance: A Survey and an Experimental Study. In: Proceedings of the 8th Annual International Conference on Industrial Engineering – Theory, Applications and Practice, Las Vegas, Nevada, USA (2003)
26. Sonderegger, A., Sauer, J.: The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. Applied Ergonomics 41, 403–410 (2010)
27. Tractinsky, N., Cokhavi, A., Kirschenbaum, M., Sharfi, T.: Evaluating the consistency of immediate aesthetic perceptions of web pages. International Journal of Human-Computer Studies 64(11), 1071–1083 (2006)
28. Tractinsky, N., Shoval-Katz, A., Ikar, D.: What is beautiful is usable. Interacting with Computers 13, 127–145 (200)
29. Tractinsky, N.: Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: Pemberton, S. (ed.) Proceedings of the 1997 Conference on Human Factors in Computing Systems (CHI 1997). ACM Press, New York (1997)
30. Tuch, A., Bargas-Avila, J., Opwis, K.: Symmetry and Aesthetics in Website Design: It's a man's Business. Computers in Human Behavior 26, 1831–1837 (2010)
31. Tuch, A.N., Bargas-Avila, J.A., Opwis, K., Wilhelm, F.H.: Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. International Journal of Human-Computer Studies 67, 703–715 (2009)
32. Tullis, T.S.: The formatting of alphanumeric displays: a review and analysis. Human Factors 25(6), 657–682 (1983)
33. Zain, J., Tey, M., Goh, Y.: Probing a Self-Developed Aesthetics Measurement Application (SDA) in Measuring Aesthetics of Mandarin Learning Web Page Interfaces. IJCSNS International Journal of Computer Science and Network Security 8(1) (2008)

Interaction Design Teaching Method Design

Chen-Wei Chiang and Kiyoshi Tomimatsu

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

Abstract. Recently, Interaction Design became popular. We found that, if we want to improve interaction design, an interaction design education method is indispensible. Especially, interaction design usually consists of two or more different fields of discipline. When members with different backgrounds are working together, there are always cause many problems, mainly during brainstorming processes. In the final, we tried to find out the main cause of the problem and tried to figure out a method for interaction design education. In the end, we also propose the possible solution to improve interaction design teaching design. We except this research can help to improve interaction design education.

Keywords: Interaction Design, workshop.

1 Introduction

Based on past records of conducting interaction design workshops, we identify three main problems that are commonly encountered by the group members, i.e.: (1) Different professional backgrounds; each member needed to spend more time to understand each other's profession. (2) Members were inefficient in developing ideas and brainstorming. (3) The results of the workshops were not as expected because of repeated discussions and readjusted ideas. Students have made comments that they would like to learn new technologies or new design methods, but due to lack of time, discussions among members had to be shortened and created unexpected results.

This research will analyze the problems that group members encounter during the interaction workshop brainstorming, such as spending most of their time developing and discussing ideas. This research is to create a teaching method to help students achieve better results and be more efficient at brainstorming.

2 Background

2.1 Related Work

There are several of researches, projects, and courses related to Interaction Design teaching method which are aimed to the students without former knowledge of Computer Science.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 45–53, 2011.
© Springer-Verlag Berlin Heidelberg 2011

In an education-friendly construction platform for wearable computing project, Ngai et al. [1] presented a programmable platform TeeBoard for wearable computing. This TeeBoard is used for training students to integrate their knowledge with physical computing.

Miyata et al.[2] have designed an education method for VR content creation for students. In this course, the education is aimed to students with the computer science background. They implied that cooperation is an indispensible element in designing contents in workshop. Concurrent thinking allows students to share their ideas through group discussion.

Ariga et al.[3] designed the course for students who are majoring in art. A toolkit was designed for students in this course which consists of both the software and hardware. They attempted to introduce interaction design to art students by using this toolkit. They designed two different versions of toolkit for their course. The second version is an improved version of the first one. Using this version, it was much easier to use sensors for students to create original interaction arts and designs.

In this course, Ariga et al. divided the toolkit's hardware into four separate elements, i.e.: I/O modules, Inputs, Outputs, and Wires. Compared to the first version, the second version of the toolkit reduced some electronic parts of Input and Output elements. Furthermore, they mentioned that the simplification setting could also aid students to understand the input and output processes.

Embodied and tangible interaction design teaching method has been discussed by Klemmer et. al.[4] and Kenneth[5]. Klemmer et. al. developed a Human-Computer Interaction Design Studio course, which is aimed to undergraduates students with HCI background, but with limited exposure to practice. The course was structured in four projects and each project is aimed to teach different aspect of embodied computing. By applying this diverse approach, the students will be able to incorporate their understanding and develop their interaction designing skills further.

Furthermore, by the classification and taxonomy of embodied and tangible interfaces was discussed by Kenneth, providing a clear guideline for the taxonomy of tangible interfaces. By applying metaphors to embodied interfaces, embodied interfaces can be plotted visually for different tasks. This taxonomy would acts as a useful tool for educators to organize their concepts before put them into real practices.

2.2 Previous Research

In Computational Thinking[6], Jeannette Wing implied that computer science education could be applied to different fields. Furthermore, Guzdial's[7] mentioned that educators need to make Computational Thinking (CT) to be accessible for everyone. Further considerations should be taken when we consider about how to introduce Interaction Design to students without computer science background. Furthermore, we also used the key questions identified by Guzdial as the considerations for our course design, i.e.:

1. What do non-computing students understand about computing?
2. What kind of challenges will they find?
3. What kinds of tools can make them access CT easily?
4. How should we, as educators organize and structure our courses to make computing accessible to the broad range of students?

Recent report in 2009[8] by Computer Science & Engineering (CISE) division of the National Science Foundation (NSF) mentioned the importance of CT. In fact, CT has been widely acknowledged not only in the field of Computer Science but also in other disciplines. The author proposed a CT based creative multi-media design educational course. The basic knowledge in multi-media education includes the knowledge of graphic, sound, and video. In this course designing, a new educating process is proposed in multi-media education, which includes technical skills, physical computing, and concept generation.

In this paper, we emphasize on physical computing education, i.e.: teaching the students on how to apply physical computing on their work. As for the final output for this education method, students will be able to integrate physical computing techniques with their design.

The definition of workshop

In general, workshops are teaching methods that divide students in groups for a short period of time. The interaction design teaching method can help students with different backgrounds develop new concepts in group discussions to create final projects. In this research we define an interaction workshop as a workshop that meets these three conditions below:

1. Continuous four days workshop.
2. Students must have computer science and media design background.
3. Every group must have a final interaction design project finished upon completion of the workshop.

The Methods of Past Interaction Design Workshops

From analyzing the information, we can conclude some important aspects of interaction design workshops and design the basic syllabus. This is to see if the syllabus can improve students at the first stage of the workshop, so they can be more efficient at developing concepts. From the documents of past workshops, we can conclude the aspects which are described below:

1. Subject: We would use subjects from one or two different areas to integrate into interaction design.
2. Analyzing students' backgrounds: Students background and their professions were analyzed, we divided them into groups so that they can exchange their information to the members that have different backgrounds.
3. Projects presentation: The end results of the workshops were functional prototypes. They were interaction design products or interaction science art installations.

Based on these aspects, we will design the syllabus and set up the basic rules to the students in our experiments.

3 Workshop Course

3.1 Workshop-1: Media Design and Physical Computing

The Introduction to Workshop

In this phase, we introduce basic physical computing tools to the workshop members. The goal is to combines physical computing technology with their professional competence, in addition to improving their interaction design skills.

Workshop Equipment Setting

In our course, we decided to use Arduino as our physical computing tool. Arduino is an open source physical computing tool, thus we could access many source code and information about Arduino and applied it to the workshop with ease. The participants were not accustomed to physical computing because they are majoring in multi-media design.

Tools and Equipment

The tools that we used in this course were consists of a hardware and a software. We used arduino as the hardware and Max/Msp as the software.

Arduino is a well-known open source tool for physical computing which means there are many references on internet. There are also many sensor modules that can be easily found on the internet. Otherwise, the students can develop it themselves by studying these references. Students can also learn it effectively in a short time, and apply those what they have learned into their works.

Max/Msp is the graphical program software where people can edit program by moving and combing some blocks than make programs come true. For beginners, this software is good for understanding how the program working. After understand how the program working, they can make their ideas come true by using this tool easily. Let their concepts or prototypes to be much more completely.

Prototyping

We divided the participants into groups and each group consists of 4-5 members. We also decided the final expected outcome for the course. In this course, our members have to think about their original concepts and finally they should make a prototype based on their original concept. The participants were expected to finish the prototype at the end of the course.

Fig. 1. (a) Wearable clothing combine with physical computing. (b) Animation controller combine with physical computing. (c) Arduino group studying.

Conclusion
Considering that the students were inexperienced in physical computing, teaching them to create a working prototype is more important than developing physical computing skills.

As the final outcome, most of the participants were able to make a working prototype based on their initial ideas. We conclude that they were able combine their concepts successfully with physical computing. However, we found that our members were giving up good concepts during the brainstorming. They thought that because of their limited physical computing and program design they couldn't realize their concepts into working prototypes.

3.2 Workshop-2: Choreography and Physical Computing

The Introduction to Workshop
In this course, we introduce Choreography of interaction to our workshop members. This concept is from Sietske Klooster et al. [9] as they designed movements and introduced them to the design of interaction, developing both as the integrated parts. In our course, we tried to let our members can understand the elements of dance and apply these element into their works for improving their interaction concepts. These concepts must come from the elements of dance and body movements.

Workshop Setting
According to Sietske Klooster et al. [9], a professional dancer is required to introduce the basic elements of dance for our workshop members. From the dance course, we identified four main elements, i.e.: Body, Space, Time, and Effort. By utilizing these elements, we had our workshop members to consider their interaction works concepts.

Equipment and Tools
Similar to the first workshop, we still used Arduino and some simple switch for the participants as the main hardware. We assumed that we could be able to reduce the gap of physical computing between the participants who have no physical computing background or experiences in the past.

Prototyping
We divided the participants into groups and each group consists of 4-5 members. We also decided the final expected outcome for the course. In this course, our members have to think about their original concepts and finally they should make a prototype based on their original concept. The participants were expected to finish the prototype at the end of the course.

Conclusion
In this workshop, we focused on how well the participants apply the dancing elements to their interaction works. As the evaluation criteria, all of the four elements should be applied on their final works. At the final presentation, the participants could submit several original works which incorporates dancing elements into interaction design.

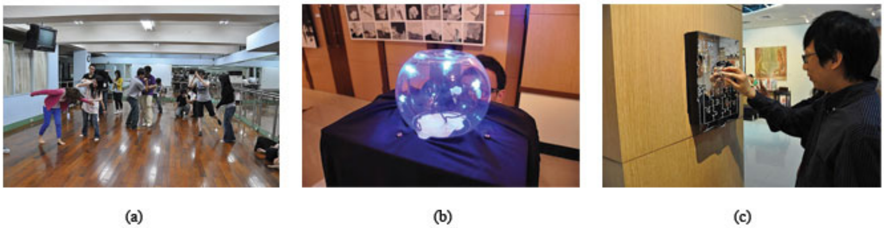


Fig. 2. (a) Dance course. (b) Work consider about space. (c) Work consider about effort.

3.3 Workshop-3: Resonance and Computing

The Introduction to Workshop

The theme of this workshop is about resonating elements. We attempted to have the participants observing the structure of resonation, applying their resonation experiences into interaction design works. In this workshop, the main objective is resonating and we defined three sub-objectives, i.e.: feeling, thinking and activating.

Workshop Setting:

In this section, the details of three sub-objective targets, i.e.: feeling, thinking and activating will be discussed in detail:

1. Feeling: using physical computing to represent peoples feeling.
2. Thinking: using wearable clothing to realize interaction design concepts. We consider wearable clothing is like human skin, thus we can represent humans feeling by using wearable clothing.
3. Activating: social network is not only exists in human society, it also exists in resonating. We tried to introduce these communicating behaviors for our members. Trying to discover what we haven't noticed sociality behavior in resonating.

In this course, we tried to let our workshop members compare resonating elements with technological elements. After the comparison, we tried to let our members to simulate the resonating elements by using technological elements and applying these elements into their interaction works.

Equipment and Tools

We used physical computing to simulate resonating environment. We used wearable clothing to represent humans feeling. As for the activating elements, we tried to have the participants simulate society behaviors.

Prototyping

We divided the participants into groups and each group consists of 4-5 members. We also decided the final expected outcome for the course in this course, our members have to think about their original concepts and finally they should make a prototype based on their original concept. The participants were expected to finish the prototype at the end of the course.

(a) (b) (c) (d)

Fig. 3. (a) Traffic monitor tree. (b) Dance movement with clothing. (c) Body movement with ball. (d) Stick a post on someone's back and get response.

Conclusion

In the final, we let our members to combine resonating elements with technology elements to create interaction works. In this workshop, we can understand, combining different fields into interaction design could create and help us to consider about that we haven't think about.

4 Lessons Learned

4.1 The Method of Past Interaction Design Workshops

During the concept development and brainstorming processes, we wrote down our ideas and discussions. In the discussion process, the main subjects changed drastically and diverge the main conversation into unrelated subjects. Sometimes group members would use post-it notes to write down ideas, but it still had the chance of making the main discussion unclear. Members would spend more time on discussions and readjust their ideas and might have to redo the concept developing process. These made the workshops inefficient and subjects could not move to the next level of the syllabus. This research will focus on improving discussion efficiency of effective group discussions.

4.2 Consider a Method for Improving Interaction Design Workshops

The training hours in a workshop usually are not long. Therefore, if we want to make workshop to be more effective, improving discussion effectiveness can be a helpful method for reducing time waste and let members have enough time for working and improve the quality of final presentation.

5 Education Method Design for Workshop

5.1 Education Method Design

Because interaction design involves various areas of knowledge, we will need to use some special terms that are not easy to understand. As the discussion progresses, it might become aimless and lower both the discussion efficiency and learning effectiveness. We propose the following methods to improve the process of the interaction workshops so that students can discuss more efficiently, reach the goals quickly, and have better final projects. The methods are:

1. Select keywords to analyze goal elements. For example, if the goal is to make an interaction design in a classroom environment, students would need to describe the space in words and record that information, such as: blackboards, chalks, chairs, desks, etc.
2. Students should record the possible ways of interacting with those objects, such as: touch, press, push, etc.
3. Analyze what computer technologies are suitable for the project and record that information, such as: light sensor, pressure sensor, temperature sensor, etc.
4. Gather the term, randomly pick one from each group, combine the words, and discuss the feasibility, such as how chairs, push, and pressure sensor can be combined to affect interaction design.

Following these steps to start discussions will help to shorten discussion time and improve the quality of final projects.

5.2 Applying to Workshop

According this interaction design education method, we applied this method to a group work concept brainstorming. In the final, we got response from workshop members. The response as the following:

1. This method for brainstorming is a very interesting method. It can help us to think about something we never think about. But, if we can have more information about these keywords, it could be more helpful for us.
2. This is an interesting thinking method for making concepts. It let me to think about what I never think about content, and I think this thinking method can help me to improve my creativity.
3. I always think my thinking method is very hard to change, but after I learned this method, I can think something I have never think very easily. I think this method can help me reach my goal of design concept much more easily. It is a very important method for me.

6 Conclusion

The process of discussion requires a lot of recording. In the future, this method can be applied as an iPad application for this research. The iPad can provide mobility to students, create and store databases easily for future references. Furthermore, This application can record the terms, then randomly select and combine terms for group discussions.

References

1. Ngai, G., Chan, S.C.F., Cheung, J.C.Y., Lau, W.W.Y.: An Education-Friendly Construction Platform for Wearable Computing. In: CHI EA 2009: Proceedings of the 57th International Conference Extended, Abstracts on Human Factors in Computing Systems, Boston, USA, pp. 3235–3240 (2009)
2. Miyata, K., Umemoto, K., Higuchi, T.: An Education Method for VR Content Creation using Groupwork. In: SIGGRAPH Asia 2009, Yokohama Japan (2009)
3. Ariga, T., Mari, K.: Learning Course for Sensory Interaction. In: SIGGRAPH Asia 2009, Yokohama Japan (2009)
4. Klemmer, S.R., Verplank, B., Ju, W.: Teaching Embodied Interaction Design Practice. Journal of Computing Sciences in Colleges 25 (2010)
5. Fishkin, K.P.: A Taxonomy for and Analysis of Tangible Interfaces. Personal and Ubiquitous Computing 8, 347–358 (2004)
6. Wing, J.M.: Computational Thinking. Communications of The ACM 49(3) (2006)
7. Guzdial, M.: Paving the Way for Computational Thinking. Communications of the ACM 51(8), 25–27 (2008)
8. Dann, W., Cooper, S.: Alice 3: Concrete to Abstract. Communications of the ACM 52(8), 27–29 (2009)
9. Klooster, S., Overbeeke, C.J., Sharp, R.: Designing Products as An Integral Part of Choreography of Interaction 2005: The Product's Form As An Integral Part of Movement. In: Proceedings of Design and Semantics of Form and Movement, 1st International Workshop on Design and Semantics of Form and Movement, Newcastle UK, pp. 23–35 (2005)

Designing Interaction Concepts, Managing Customer Expectation and Mastering Agile Development in Rich Application Product Development

Marcela Esteves and Vladimir Andrade

Siemens Corporate Research
755 College Road East, Princeton, NJ 08540, USA
{marcela.esteves,vladimir.andrade.ext}@siemens.com

Abstract. The emergence of rich application implementation frameworks, such as WPF and Silverlight, promoted a new collaboration paradigm between developers and designers where ownership of the user interface code is transferred to the user experience team. The implications of this new paradigm for the user centered design process impact its technical, collaborative, and business dimensions. The traditional design prototype can now demonstrate most of the desired user experience and could be directly integrated with the back-end code, significantly reducing the design revision costs. Creating the rich prototypes demand enhanced technical skills from visual designers, who become a member of both the design and implementation teams. The implementation tools provided by the rich application frameworks aim to simplify the prototype creation task for the designer, but can potentially lead customers to expect a lower effort associated with the user centered design process.

Keywords: user centered design process, WPF, Silverlight, visual design, interaction design, rich application.

1 Introduction

User Centered Design (UCD) describes a multidisciplinary user interface[1] development process that positions the user's goals, workflows, and requirements as the main drivers of an application's navigation, interaction, functionality, and behavior. UCD promotes early user involvement in requirements definition and user interface evaluations, which should be performed in multiple iterations until a final solution is reached. The UCD process has been advocated and practiced by multiple members of the user experience community with successful measurable results for the past decade and is supported by an array of established methodologies [1,2,3]. The international standard ISO 13407 [4], "Human Centered Design Process for Interactive Systems", identifies four activities that are part of a best practice process: specify the context of use, specify user and organizational requirements, produce design solutions, and evaluate designs against user requirements.

[1] We focus the scope of "user interface" on software interfaces only.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 54–62, 2011.
© Springer-Verlag Berlin Heidelberg 2011

The context of usage analysis involves understanding the users' profiles, their goals, tasks, workflows, decision making patterns, and environments. This knowledge is consolidated in a user behavior model, which allows the user experience team members to build a solid framework that will later enable the definition of a user interface concept - including information architecture, navigation, and interaction - aligned with the users' mental model. Contextual inquiry, focus groups, personas/user profiles are examples of methodologies used in context of usage analysis.

Requirements define the functional scope of an application across multiple versions, focusing on what is needed by different stakeholders ("problem space") but not on how those needs should be addressed in the user interface ("solution space"). While the context of usage and requirements specifications are sequential activities in the process, it is not uncommon within the industry for the requirements to be defined prior to or in parallel to context of usage analysis. In these cases, the user behavior modeling provides usage context for the requirements and helps define the focus of each application version based on the users' priorities. Use Cases have proved to be a useful tool to connect user requirements with user interface concept: Use Cases contain a detailed task description, and tasks are connected to user needs which are then documented as requirements.

The design solution and design evaluation steps of the UCD process are performed sequentially over multiple iterations where early user feedback guides the user interface towards a useful, usable, and enjoyable application. Design solutions start as high level user interface concept alternatives (e.g., documented as wireframes) focusing on basic information architecture, navigation, and interaction models, and mature over multiple iterations to become the final complete solution that addresses all requirements (e.g., final design screens covering all steps for task completion). User feedback can be gathered through different validation techniques using low or high fidelity prototypes depending on the maturity level of the user interface solution.

Among UCD's strengths, its ability to provide a user interface solution that is validated by the user is of special interest for a fruitful collaboration between user experience and software development teams: when applied early in the product life-cycle, UCD process has the power to inform the implementation team and significantly decrease development effort [4]. Over the past years UCD has been combined with frequently used software development models, the V model and Agile development.

The V model process organizes the deliverables that are part of software development along product definition and product testing steps that correspond to specific validation and integration levels. Requirements, Functional, and Design Specs are sequentially created as part of product definition and serve as a roadmap for the Unit, Sub-system, and System tests that take place prior to commercialization. A smooth integration of the UCD and the V model relies on 1. the context of usage and requirements definition activities taking place prior to the Requirement Specification milestone; and on 2. the design solution and evaluations to be completed for high level user interface concept and mostly complete for detailed requirement screen design before the Design Specs milestone, when implementation starts.

Agile development [5] is a feature centric process where requirements are grouped in smaller sets that describe a user story. The implementation team aims to deliver those sets in quick iterations ("sprints") that allow the customer to validate working software early and multiple times during the development process. Research on the topic of combining UCD and Agile development [6, 7, 8] has indicated that the most successful approach is to allow sufficient time for UCD's context of usage analysis, requirements definition, and high level design solution activities to take place prior to implementation. In this approach the high level user concept design is previously defined and screen design refinements needed to address specific user story points are included in the sprint planning activity of Agile development.

During the aforementioned software development models, the prototypes that allow early validations with users are created using technology and tools that require lower implementation effort and can be confidently utilized by resources outside of the development team. Examples include Microsoft's Power Point, Axure RP, and Adobe Flash platform. The "work in progress" nature of the iterative process prevents the prototypes from being integrated into the application's code base, since early in the UCD process the user interface concept suffers dramatic changes with each iteration[2]. As a result, once the application's navigation, interaction, and behavior are stabilized, the prototype becomes a reference to the development team, who then has to re-implement the application user experience in a different code platform.

However, a new paradigm has emerged where the difference between prototype and application code is blurred, with important implications for the UCD process as discussed below.

2 Rich Applications and the Emergence of a New Software Development Paradigm

A Rich Internet Application (RIA) is an application that normally runs on a web browser, while providing similar functionality to traditional desktop software without needing to be installed into the user's computer. These are developed either using cross-browser and cross-platform technologies, such as HTML, JavaScript and CSS or using application frameworks and tools developed by third party companies, like Microsoft's Expression Blend, Adobe's Flash /Flex technologies and Sun's Java.

These technologies, when compared to regular, static web pages, enable a richer user experience by giving the software creators control over interactions, with custom and more engaging controls that improve overall user experience, system and network performance, with localized data refreshing that reduce server requests, content reloads and data transfers, while also providing opportunities to build a strong visual language.

With the advent of Web 2.0, social media and user-generated content, RIAs can be found everywhere: webmail clients, content dashboards /start pages, blogs, social networks, video sharing websites -- they are an integral part of today's internet culture. And, with the rising popularity of connected mobile devices, they are available virtually everywhere with little to no functionality loss in their mobile counterparts. Due to the broader appeal of the Internet, RIAs began setting the standard in terms of

[2] While Agile development allows for code validation over multiple iterations/sprints, it doesn't accommodate significant ongoing changes to the user interface concept due to its impact on the software architecture. Therefore, the user validations that focus on the basic user interface concept are performed prior to implementation using prototypes.

visual design, its supporting elements and how the user interacts with the software. This, in turn, increased the user's expectations towards traditional desktop applications. They expected lighter, more usable, better-looking and more interactive software, similar to what they could easily find in their web-browser.

While RIAs were wildly popular, they also facilitated the creation of a new development paradigm. While the standard model, in which a designer sends design files with specifications to the development team, is effective, it is also very time consuming: for every design file sent, rounds of validations and corrections are required to achieve acceptable implementation results. Because of the shorter turnaround times typically expected when developing for the Internet, it became important to develop ways of shortening the amount of time and effort spent on user interface implementation and validations.

Within the new paradigm, the designer is also a part of the implementation team. He is the one responsible for turning his own design files into usable front-end code. This is made possible by tools, such as Microsoft Expression Blend and Adobe Flash Professional, which facilitate the generation of said code by providing a graphical user interface that is somewhat similar to the tools traditionally used by designers.

While the front-end code, once integrated with the back-end, is an integral part of the project, it is also a completely independent entity works and run on its own as a high fidelity prototype, with static content and functionality. With regards to the UCD process, working on a high-fidelity prototype from the beginning of the project could be very time consuming. Therefore, low-effort prototyping for early user validations might be better suited for the early iterations in the project. Then, once the visual language is defined, assisted by design tools such as Adobe Photoshop and Illustrator, the designer can proceed to create a high fidelity prototype with the final design specifications. This, however, is a two-step process, as design tools are better suited for design tasks as discussed below.

Following this process enables great savings in effort and time, since the development team is freed from the task of translating design specifications into front-end, thus focusing more on the actual back-end coding. This also reduces the amount of rounds needed to refine the front-end implementation since this task is now in the hands of the designer. Now, the developers integrate, instead of implementing the front-end.

3 UCD Process and the New Software Development Paradigm

The new software development paradigm allows a more efficient collaboration between designers and developers, leading to an optimized implementation process with lower costs and decreased time to market. This optimization can only take place to its full extent if the user experience team, with special regards to visual designers, embraces the ownership not only of the user interface definition but also of its implementation. To successfully bridge UCD and the new development paradigm, the visual designer needs to acquire additional technical expertise, maintain a clear direction with regards to the application's visual language, and understand the development team's needs from an implementation perspective. Furthermore, the user experience team should educate and manage customers' expectations regarding the impact of the user new interface code activity on the standard UCD process.

3.1 Visual Designers' Skill Set

Visual designers define the application's visual language through a consistent graphic representation of all user interface elements that promotes the navigation and interaction models. In performing this systematic and creative task, visual designers are concerned with hierarchy, relationship, and weight of visual widgets to harmonically combine them in a high fidelity screen design. Designers typically use dedicated tools that provide extensive functionality to create and manipulate widgets, with Adobe Illustrator and Adobe Photoshop being among the most prominent examples.

By contrast, implementing a visual design requires a completely different mental model from defining visual design. The high fidelity screens provided by the visual designers are decomposed in connected, coded controls associated with pre-defined behaviors and states. While visual design definition is a conceptual task, visual design implementation is an execution task. In the new paradigm, the visual design implementation is owned by the user experience team, specifically the visual designer, which allows the front-end and back-end coding activities to take place in parallel and ensures the application's user interface matches the design specs.

While some rich application platforms offer dedicated tools that are designer friendly, such as Microsoft Expression Blend, those tools are not organized around a conceptual workflow but rather around an execution workflow. For instance, on design tools graphical shapes can be combined to create the visual properties of a UI element such as a menu, without the need to consider what type of control the menu may be when the screen design is implemented. Meanwhile implementation tools offer controls that have to be nested into a larger screen structure, so there has to be an a priori choice on the type of control to be used (e.g., a menu could be a grid or a stack panel) so its graphic properties can then be manipulated. Furthermore, despite the availability of implementation tools, full control over interaction behavior typically is only accessible through code writing. Therefore, the benefits brought by the new paradigm depend on the visual designer's ability to learn the implementation mental model and related technical skills.

3.2 Collaboration

The new development paradigm promotes intense collaboration between designers and developers during the implementation process. Whereas in traditional software development models designers hand over specifications to the development team and later provide input on design corrections needed on the implemented code, in the new paradigm designers deliver front-end code and are part of the development team. As designers and developers create a partnership over the same code case, the quality of the application's user interface code dramatically improves since designers make sure that the front-end implementation is fully compliant with the user experience specifications they created.

The collaboration process can be greatly improved if a developer in an "integrator" role is available to support the visual designer during implementation. The integrator not only connects the front-end code delivered by the designer with the back-end code, but also supports the designer with implementing complex and/or non-standard user interface behaviors that require heavy code writing.

3.3 Visual and Interaction Design Bias

While not a direct consequence of the new development paradigm, it is worth noting that visual designers need to be attentive to avoid bias towards the look & feel (e.g., gradients, 3D effects) of standard controls provided by the implementation tools such as Expression Blend.

The history of software development shows that visual languages established by large software vendors such as Microsoft and Apple for a given operating system release tend to influence the look & feel of applications developed during the release's life cycle. Implementation tools reflect the current look & feel and as such will likely be updated when a new operating system is launched. Because consumer oriented applications are typically enhanced through frequent releases, they can more easily adapt to changes in the industry's look & feel status quo. However, visual designers working on business to business industries should consider that the typically long life cycle of these industries' applications requires their visual language to persist through look & feel trends to avoid being outdated and alienating users.

3.4 Customer Expectation

The implementation tools available in some of the rich application platforms substantially decrease the effort needed for a visual designer to create functional front-end code. Customers of user experience teams can be led to believe that the simplification provided by the designer friendly implementation tools will cause the design solution activity of the UCD process to require less effort since elements of the user experience are promptly available. In that case the user experience team should aim to continuously educate customers about the scope of the design solution activity which entails developing a user interface concept that defines navigation, interaction, and behavior in contrast with the widget level control provided by the implementation tools. Developing applications using rich platforms and their implementation tools provide for appealing widget behavior but doesn't guarantee a useful, usable, and enjoyable application.

4 Examples of Rich Application Product Definition in Agile Development

Two previous and current product development projects of rich applications developed in Microsoft's WPF and Silverlight provided a body of experience in successful execution of UCD process in the new paradigm within the agile development model.

The first project concerned a platform development in WPF that took place between 2008 and 2009, included approximately 2,500 market requirements for the first product release, and involved 10 development teams distributed in 5 locations over 3 continents. The UCD process was followed from project start and the basic user interface concept containing navigation, interaction, and behavior definition was completed within the first 8 months of the project. During that period the platform's visual language was also defined. Each application provided by the platform was divided into a module that was implemented during one or more sprints. While the basic user interface concept and visual language was defined, the design solution of each module had to be refined to account for interaction model to support detailed requirements and the visual design of specific UI elements. Those two activities were performed for each module as part of the project's sprint schedule.

Figure 1 shows how the interaction and visual design refinement for each module was included in the sprint schedule. The interaction and visual design refinement tasks for the first module to be implemented ("Module A") and the interaction refinement for the second module ("Module B") took place before sprint 1. During sprint 1 and subsequent sprints, front-end code, interaction refinement, and visual design refinement happened in parallel for different modules, to ensure that when the sprint for the implementation of a given module started, this module's interaction and visual design were finalized and approved.

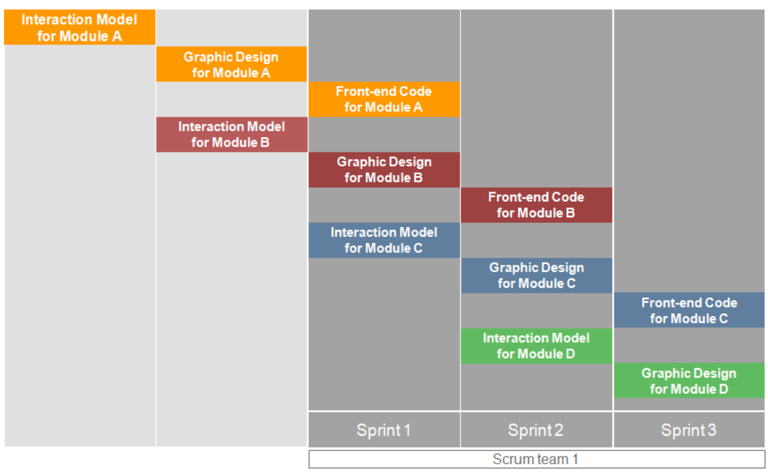


Fig. 1. Sprint Schedule accounting for UCD process activities and front-end code delivery by user experience team

During that period the Expression Blend tool provided by the WPF platform was in its infancy (e.g., during the project duration the visual designers worked with Blend's beta version and release candidates). Given the high complexity of the software architecture, the fact that both user experience and development teams had to learn the new WPF platform, and the state of the Blend tool, the collaboration paradigm chosen was as follows: during the sprint the development team would deliver to the user experience team the module's functional code, and the visual designer's main role was to style it in the Blend tool/code with continuous support from the scrum team's integrator. In this project the visual designer didn't create the front-end code from scratch.

The second project is a product development based on an existing platform that is currently coded in WinForms. This project is developed in Silverlight, started on May

2010 with first product release by end of 2011, and involves 5 development teams distributed across 2 continents. The UCD process was also followed from project start with the design solution activity completed prior to sprint implementation. In the course of 6 months the user interface concept, interaction refinements to address all requirements, and visual design definition were completed in an iterative process involving user validations. The product is also divided in modules which are developed over one or more sprints.

The visual designer's role within development is to create the front-end code from scratch for each module and to create a Resources Dictionary (library of UI components' code supported by the xaml programming language that is behind the WPF and Silverlight platform) for the entire application. Because this project's software architecture complexity is more easily managed and because the Silverlight platform and Blend tool have greatly improved over the past couple of years, the visual designer is able to create a module's front-end code in Blend and deliver it to the development team who then integrates the module's front-end and back-end. In addition the visual designer also saves all UI components (called "resources" within the Blend tool, primarily composed of styles, templates, and brushes) into the Resource Library which is used to manage the look & feel of all modules.

Because of the aforementioned difference between visual design definition being a conceptual task in contrast with visual design implementation being an execution task, the visual designer first created the module's screen designs in Adobe Photoshop (which provides better control over visual properties than Adobe Illustrator) and then recreated the module in Blend to generate the front-end code. In this process the visual designer is supported by a senior developer who troubleshoots resolves front-end code behavior needed for given modules.

5 Conclusion

The new software development paradigm is a major step forward towards resolving the previously discussed inefficiencies caused by the hand-over of the user experience that happens between designers and developers in the traditional collaboration paradigm. The main challenge for user experience practitioners, with special regards to visual designers, is learning a new, complex skill set that typically lies beyond the interest of visual design domain experts.

Furthermore, most of the rich application platforms such as Adobe Flash learned by students during their academic interaction and visual design courses are not typically associated with product development. This creates a situation where visual designers hired by companied that build desktop applications are typically not familiar with rich application platforms that are beneficial for product development, such as WPF and Silverlight. This causes the ramp-up phase and learning curve of visual designers to be time and effort consuming when they are first exposed to those platforms.

The currently available designer friendly development tools are widget centric and tend to require that visual designers first approach the front-end implementation and then define the visual design. However, the definition process often involves experimentation with different approaches to the UI elements that may require revisions to the interaction concept and to the types of controls used.

Ultimately, the design definition and implementation processes are different in nature and will likely continue to be performed as sequential steps, in the same way that an architecture plan is defined prior to construction. The ideal situation for visual designers would be a tool that allows the screen designs that emerge from the conceptual design definition task to be easily converted into front-end code. This approach would eliminate the need from visual designers to either perform the design task in a development tool or to recreate the design into a software development tool.

References

1. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann, San Francisco (1997)
2. Cooper, A., Reimann, R., Cronin, D.: About Face 3: The Essentials of Interaction Design. Wiley, Hoboken (2007)
3. Rubin, J., Chisnell, D., Spool, J.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. Wiley, Hoboken (2008)
4. International Organization for Standardization, http://www.iso.org
5. Martin, R.: Agile Software Development, Principles, Patterns, and Practices. Prentice Hall, Upper Saddle River (2002)
6. Fox, D., Sillito, J., Maurer, F.: Agile Methods and User-Centered Design: How These Two Methodologies Are Being Successfully Integrated in Industry. In: Agile 2008 Conference (2008)
7. Labib, C., Hasanein, E., Hegazy, O.: Early Development of Graphical User Interface (GUI) in Agile Methodologies. J. Comp. Met. Sci. Eng. 9, 239–249 (2009)
8. Patton, J.: Designing Requirements: Incorporating User-Centered Design into an Agile SW Development Process. In: XP/Agile Universe 2002. Springer, Heidelberg (2002)

POLVO - Software for Prototyping of Low-Fidelity Interfaces in Agile Development

Júnia Gonçalves and Caroline Santos

Computing Departament,
Universidade Federal dos Vales do Jequitinhonha e Mucuri,
Diamantina, Minas Gerais, Brazil
{junia.goncalves,carol.qs}@gmail.com

Abstract. In the process of software development, the ease with which the user can perform his/her tasks in the system - commonly called usability - is an important requirement. The prototyping of user interfaces is one of the most widely used techniques to specify this type of requirement. This paper presents the importance and need to improve and increase the agility of prototyping interfaces in agile development processes. The authors propose a software that is able to build low-fidelity prototypes, document them and support user testing, to aid the process of interface building in the Scrum methodology.

Keywords: Low-fidelity Prototype, Interaction Design, Agile Methods, Scrum.

1 Introduction

Traditional models of software development, such as Waterfall, are generally focused on following the plan, instead of satisfying the customer. So when changes are needed, either requirements or technology, this results in an increase in costs proportional to the stage where the project is [6] - The more advanced the process, the greater the increase in costs. Even the unified processes of software development following the incremental-iterative[1] model are not focused on customer satisfaction. Thus, agile methods have gained importance in software development, rather than traditional processes, since the former are better suited to dynamic environments and tight deadlines, [6]. This adaptability exists because the agile methodologies are focused on continued delivery of a functioning system, the reduction of the burden of documentation on the development and frequent contact with the client, resulting in improved response to changes [5]. Among the agile, Scrum stands out with its product management practice.

For applications to stand out today, it is necessary that, for their development, a stance toward the user be adopted, with interfaces that are usable and easy to learn, because when they are not, more resources will have to be used in training and

[1] According to Sommerville [17], the iterative-incremental models are those in which the system requirements are identified and prioritized, followed by a series of development stages, and each of these stages results in the delivery of a subset of system functionalities.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 63–71, 2011.
© Springer-Verlag Berlin Heidelberg 2011

supporting the user, and this creates dissatisfaction with the tasks to be performed [1]. However, interaction design approaches are not clearly provided in software development, either in traditional processes, or in agile methodologies; therefore, the systems do not give support to usable form tasks [3]. Thus, it is necessary that an attitude directed to the user and use be inserted during the procedure in order to add greater value to the final product.

In the process of software interface design, prototypes are designed to facilitate communication between developers and stakeholders [12]. According to Ambler [2], the conceptual modeling in agile methodologies happens in draft form, drawn on white boards, sheets of paper or flip charts, as these sketches are enough for the explanation of what should be modeled and then produced. Besides, Nielsen [10] believes that the performance of processes focused on system usability can be improved through computer support, stressing that it is desirable to develop an application that supports the creation of mock-ups of user interface and user testing. Thus, this study proposes to develop software for a low-fidelity prototyping, agile practices aligned with the Scrum methodology, supporting documentation and usability testing.

This work consists of five sections, the first of which presents the problems worked on here, as well as their motivations. Section number 2 presents concepts and theories about the design of user interfaces, followed by section 3, which presents concepts on software engineering and agile methods. The subsequent sections, 4 and 5, are related to the evolution of concepts and theories mentioned above, basing the development methodology used and details of the methodology concerning the design of the application, together with the documentation of its development. Finally, the concluding remarks are presented.

2 Human-Computer Interaction and Interface Prototype

Once systems began to be marketed on a large scale and distributed to many users, also started the concern that users could interact with the system. It became necessary to adapt these systems to the needs and goals of users, facilitating the use and development of their work and leisure activities. In literature, there are three main approaches for developing system interfaces, usability engineering, user-centered design and usage-centered design. The first is a process directed to the ease of learning and use of systems, providing them with a friendly interface. The second is to change the focus of development, leaving the focus on technology, moving to focus on the user, in which he/she is studied, included and takes part in testing. The last process focuses on the use of the system, in which activities it should be conducted and what the system should cancel, leaving the figure of the user aside [14].

The User-Centered Design (UCD) is a philosophy that is based on the needs and interests of users, emphasizing the creation of usable and understandable products [11]. Thus, the UCD approach can be used to create various products, including software. As Donald Norman introduced the concept of User-Centered Design, was also embedded the concept of Activity-Centered Design, in which tasks and user behavior were studied for framing systems. But later, Norman concluded that the model-driven tasks produced inadequate results [4]. Thus, Cooper, Reimann and Cronin [4] suggested that not only the profile, the activities and environment of users be investigated, but also their goals. The technology-focused organizations rarely have a proper UCD process, if ever presented [4]. Even if the Activity-Centered Design manages a system that can be modeled, this is just one increment, since it does not provide solutions to differentiate the product on the market and does not satisfy the user correctly. Thus, the Goal-Directed Design (GDD) is to be a link between the research user with design, using ethnographic techniques, interviews with stakeholders, market research, detailed models of users, scenario-based design and a set principles and patterns of interaction, meeting the needs of users and organizations.

2.1 Interface Prototype

The prototyping of interfaces can be defined as a limited representation of a design that enables users to interact and explore it, and its main idea is to create something that resembles the final product and can be tested by end users, saving resources [10]. In this work, prototypes will be understood as limited representations and non-executable user interfaces, which can be manipulated in order to validate the actual interfaces of the system under development. Prototypes can be classified into high-fidelity prototypes and low-fidelity prototypes. The first type of prototype is one that closely resembles the final product, being developed under a programming language and sometimes executable. They are used to demonstrate a real image of the system and evaluate patterns and style guides [14]. The low-fidelity prototypes are those that have little resemblance to the final product, using simplistic means for their representation, instead of metal and electronic displays. These prototypes are useful, since they tend to be simple, cheap and rapidly produced. They can be quickly modified, supporting the exploration of alternative designs. They are useful for identifying market requirements, assess multiple design concepts, deal with issues on layout of the screen [12], and, finally, are useful for exploring the possibilities of navigation [14].

For Snyder [16], paper prototyping is a technique that involves the creation of low-fidelity prototypes on paper that can be manipulated by a facilitator that simulates the behavior of the system to conduct usability tests. This technique provides benefits in terms of skill development, ease of communication between multidisciplinary teams and stakeholders, encourages creativity, does not require technical skills, which allows end users to be involved in prototyping the interface, enhancing the quality of the final product [5], though users might find it strange at first [16].

3 Agile Software Development

Software Engineering consists of a set of activities that aim to create software product - commonly called software process [17]. In the 90s, dozens of developers joined together, sharing the dissatisfaction with the prevailing process of software development - UP[2] - and wrote the so-called Agile Manifesto, initiating a series of agile methodologies. In the context of software engineering, agility reflects the capacity to accommodate the necessary changes that originate during the process of software

[2] Unified Process is a methodology for developing iterative-incremental software, which is characterized by extensive documentation through UML diagrams [17].

development [13]. Thus, agile methods present a great response to the changes, since following the plan and providing comprehensive documentation are items of low priority, indicating the absence of a rigid structure, which leads to creativity, self-manageable processes and increasing returns. Besides agility, agile methodologies embody other values to their process, such as the creation of cohesive teams, communication between the implementation teams, engineers, managers and stakeholders, and the latter are considered part of the development team and add greater value to products that are potentially deliverable to the customer.

Scrum is a framework for agile development of complex products, within which it is possible to employ various processes and techniques [15], one method to manage the development of a product of any technology, including software [7]. The devices available in Scrum are the Product Backlog, the Sprint Backlog, the Burndown Release Delivery and the Sprint Burndown. The Product Backlog is a list that prioritizes all that is needed in the product. The Sprint Backlog is a list of tasks to turn the Product Backlog into a potentially shippable product. Both Burndown Release Delivery and Sprint Burndown charts measure how much effort and time are still needed. The former is directed to the Product Backlog and the second to Sprints. Moreover, in Scrum, there are time-boxes, ie, events of fixed duration, which are the Release Planning Meeting for Delivery and Sprint.

4 The Project

The purpose of the software presented here is to develop a system of low-fidelity prototyping, according to the story of the Scrum framework and tasks that require a user interface and supporting documentation and usability testing. The development was focused on construction of key features, with no effort at that time to implement non-functional requirements such as security and performance, because it is an early application in its evolutionary chain. However, the non-functional requirement known as usability has been strongly sought and tested. In order to obtain usability in the system, an approach that would integrate UCD and agile methods was chosen, but there is no formal model for this, which has led several authors to apply effort in the integration of UCD with models of software development with agile methodologies [8, 9, 18]. To make this integration difficult, there are still UCD activities that are not effective in practice [19], making specific case studies to be developed to test the compatibility of these UCD activities with agile methods [8].

A model for such integration is suggested by Sy [18], which is the focus of the agile team on a few features of the system at a time, which means that each feature receives a more careful work, and can capture design flaws earlier, provide changes and incorporate these adjustments. In addition, there must be different cycles of design and implementation, although they may occur in parallel, in which functionality is implemented with a basic interface as its design is made. So, the next iteration, the final design is implemented. There should also be the so-called Cycle Zero, which is the phase of gathering usability requirements.

4.1 Development Methodology

Based on Sy's model [18], a similar methodology was used in this work, consisting of a mix between Goal-Directed Design process and Scrum agile methodology. The Zero Cycle is divided into three activities of requirement gathering system based on the users' goals, according to the initial stages of Goal-Directed Design. Subsequently, the Sprints take place in parallel with the design process and prototyping of the interfaces, resulting in an interface refining activity through usability testing and participatory design sessions. Finally, the final interface can be implemented. The activities of the methodology are presented below and illustrated in picture 1.

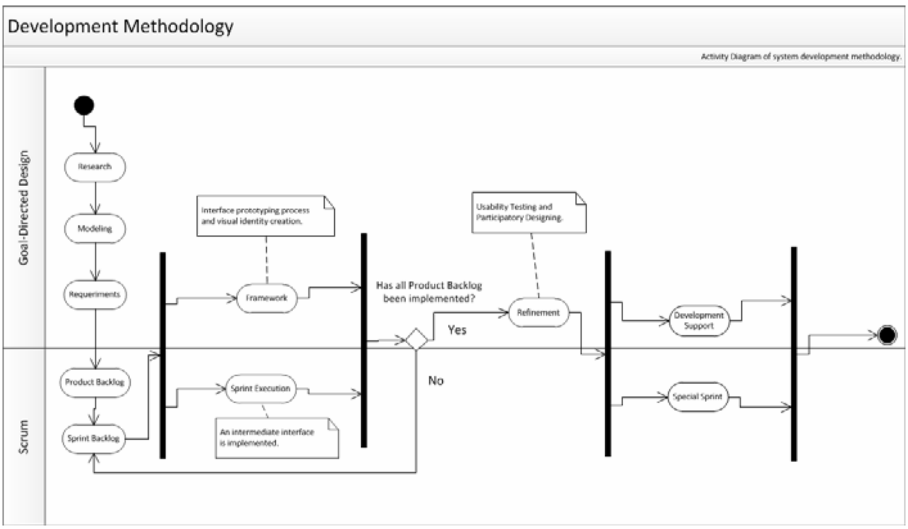


Fig. 1. Development Methodology: Mix between Goal-Directed Design and Scrum

The activities consist of Research: it is the literature review of the areas of Software Engineering and Human-Computer Interface, and is presented and cited throughout the text; Modeling: it is the creation of personas and goals based on information obtained during the research; Requirement Definition: it is the analysis of the personas and goals to generate the specification of the users' scenarios; Definition of the Product Backlog: from scenarios, The Product Backlog is built on the application, together with the story definition, its importance, demonstration, initial estimates and essential UML diagrams; Definition of Sprint Backlog: The development loop occurs in activities 5, 6 and 7, starting with the activity known as Sprint Backlog Definition, when the stories are chosen and will compose the current Sprint, which leads to a potentially shippable product until the entire Product Backlog is implemented; Defining the Framework: activity that occurs parallel with Execution of Sprint, in which occurs prototyping and creating the visual identity of the application according to the stories present in the current Sprint; Sprint's implementation: it is the implementation of selected stories in the activity known as Setting the Sprint Backlog, with the development of an intermediate interface that will be replaced later;

Refinement: it happens after the loop of activities 5, 6 and 7, consisting of performing usability testing, participatory design sessions with users of focus and redesign of user interfaces; Special Sprint: it consists of a Sprint in which end-user interfaces are implemented, replacing the interfaces developed in the intermediate activity of Sprint Implementation. This activity takes place parallel with the activity Development Support. Development Support: it consists in monitoring the implementation of end-user interfaces if any development challenges are identified and lack some adjustment.

4.2 POLVO and Its Agile Development Directed by Goals

This is the breakdown of major activities conducted during the development of the system:

System Models. The system was modeled after the creation of six personas, models that focus on users of the system focusing on characteristics and behavior [4], representing its primary users, secondary, additional, served and negative. The primary personas are those that are directly related to the functionality of the system, while the secondary personas are satisfied with the features presented, in spite of having special needs. The additional persona is one that is not primary or secondary, whose needs are a combination of the needs of primary and secondary users, and is completely satisfied with the functionality of the system. The served persona is the one who does not use the product, but is affected by its use, while the negative persona is one whose needs are not supported by the product [4]. The personas were described according to their context, and all the personas were, on some level, set in the context of software development with Scrum. They were also described as for the activities they carry out, their attitudes, skills, motivations and goals. Moreover, they were mapped as to their experiences and knowledge on technology, Human-Computer Interface, participation in the implementation of applications and interest in the prototyping process in order to show the level and the difference between the personas.

Requirement Definition, Product Backlog and Sprints. Based on established personas, their scenarios were developed, which consist of how the persona uses the product in its day-to-day activities in order to achieve its goals [4], and these scenarios are the system requirements, added to the device Product Backlog as part of user stories. Thus, the stories are constructed based on user scenarios, divided into six categories: Scrum, Prototyping, Usability Testing, Security and Documentation, with the first three being the highest priorities. According to the device Product Backlog, the software must support an authentication system and different access control to "designer" and "user", and only the "user" can perform usability tests, preventing the "designer" from performing tests and generating erroneous results, but the designer can still view the navigation between the prototypes. The "User" can also create interface prototypes, but the user cannot view or modify prototypes that are not his property. Also, it should support product management, user stories, tasks and Sprints and construction of prototypes based on the tasks to be performed in parallel. Usability testing should be performed by Sprint since the prototypes are interconnected through links or actions, and must mark what prototype is the Sprint start screen. It is important to note that the prototypes are built as interactive and non-executable, that is, are navigable, but do not process data, similar to the prototypes on paper. The present application versioning allows alternative prototypes to be created and stored, providing a documentation of the construction of interfaces, just like the notes come from the documentation generated by the designer during the usability test. Finally, the user can add "designer" or "user" to his product in order to generate collaborative development processes, although the change control is left to the users.

The Sprints were defined according to the importance of the user story in the Product Backlog, followed by the duration of the stories already chosen. Thus, three sprints were defined, the first forming the core functionality of the application, construction of prototypes, Products, Stories and Tasks and Sprints, the second consisting of Versioning features, Notes and Execution of Usability Testing, and finally the last Sprint is made up of the features of Result of Usability Testing, System Authentication and Access Control and Adding Designer / User.

Prototyping, Usability Testing and Special Sprint. Alongside the implementation of Sprint, there was the process of low-fidelity prototyping on paper in the application. Several alternative prototypes were built until one was chosen based on a Heuristic Evaluation, out of which the most usable prototype would be built on interactive paper, to be used in usability testing with users on focus. Thus, the prototypes were built on paper according to the stories implemented in the Sprint, and the interactive prototype was finally built. In addition, the visual identity of the application was also designed, being partially embedded in the interactive paper prototype and fully incorporated into the final interface implemented.

At the end of the loop in the activities of Sprint Definition, Framework Definition and Implementation of Sprint, the activity known as Refinement of the user interface started, through sessions of usability testing conducted with a small number of users on focus (from 1 to 5), who were invited to develop a product that consisted of a website for lyrics, in which these were published collaboratively by its users. This product was named "Lyrics", with four user stories, four tasks, two Sprints and two interface prototypes. The tasks of the usability testing focused on product, user stories, tasks and Sprints management, prototype building, versioning, annotations and adding other users. Participants were free to create / prototype interface as they wished. Usability testing consisted of an initial questionnaire of adherence to the focus group, the tasks to be performed on prototype interactive paper and a final questionnaire on the impressions of the software, as well as being invited to "think aloud". Based on the responses of the final questionnaire, participants were invited to a meeting of Participatory Design in which they built their solutions and / or ideas on interactive prototype paper, using simple tools such as pen and scissors. The solution was built and tested if there were some related problem detected during testing, whereas the ideas, concepts and opinions about the visual identity were not incorporated into the prototype. The biggest problems were the creation of user stories and their allocation to Sprint and adding designer / user.

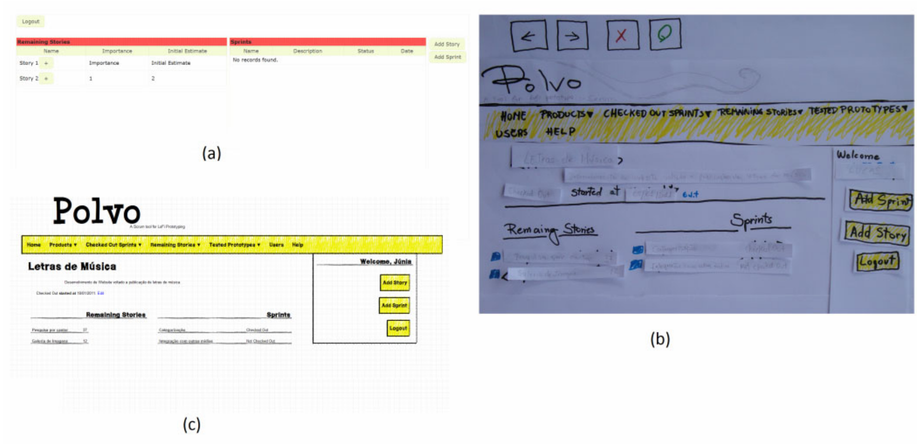


Fig. 2. Intermediate interface (a), interactive paper prototyping (b) and real final interface (c)

After the sessions of Usability Testing and Participatory Design and redesign based on these sessions, the interactive paper prototype was implemented along with its previously designed visual identity, featuring Sprint Special. During the implementation of the final interface user, development challenges were not detected, not requiring design review.

5 Conclusions

The prototyping of user interfaces is perceived in various software development methodologies, since it presents a better way to express the requirements of the user interface, instead of diagrams and textual descriptions [17]. There is also a need for these prototyped interfaces to be tested and documented. Because of that, the software "POLVO" was proposed, which brings such features. Possibly, the use of "POLVO" will bring benefits to software developers, such as increased agility in the development of user interface prototypes, ease of usability testing application with interactive prototypes, carrying out participatory design sessions and documentation of prototypes.

Although the tool has been developed, it needs to be evolved, adding other non-functional requirements such as enhanced security, performance, reliability and availability. Moreover, its functionalities can be expanded to include all of the Scrum management added to its current capabilities. The construction of such software has emerged from studies on agile development and prototyping of interfaces. For evaluation of its features, the use of the software by a team of developers would be required. Thus, within the proposed evolution of the software "POLVO", there can also be case studies focusing on the integration of the agile process Goal-Directed Design and Scrum and adaptations to the current model used.

References

1. Ambler, S. (2005-2006). The Agile Unified Process (AUP), `http://www.ambysoft.com/unifiedprocess/agileUP.html`
2. Ambler, S.: User Interface Design Tips, Techniques, and Principles, Ambysoft
3. Blomkvist, S.: User-Centred Design and Agile Development of IT Systems, M.S. thesis, Departament of Information Technology, Uppsala University, Uppsala, Sweden (2006)
4. Cooper, A., et al.: About Face 3: The Essencials of Interaction Desgin. Wiley Publishing Inc., Indianapolis (2007)
5. Gullisken, J., et al.: Key Principles for User Centred Systems Design. Behaviour and Information Tecnology 22(6), 396–409 (2003)
6. Highsmith, J., Cockburn, A.: Agile Software Development: The Business of Innovation. IEEE Computer 39(9), 120–127 (2001)
7. Koch, A.: Agile Software development: Evaluating Methods for your Organization. Artech House, Norwood (2005)
8. McInerney, P., Maurer, F.: UCD in Agile Projects: Dream Team or Odd Couple? Interactions 12(6), 19–23 (2005)
9. Memmel, T., et al.: Agile Human-Centered Software Engineering. In: People and Computer XXI HCI, Beijing, China (2007)
10. Nielsen, J.: Usability Engineering. Academic Press, Boston (1993)
11. Norman, D.: The Design of Everyday Things. Basic Books, New York (1988)
12. Preece, J., et al.: Interaction Design: Beyond Human-Computer Interaction, 2nd edn. Wiley Publishing Inc., Indianapolis (2007)
13. Pressman, R.: Software Engineering: A Practitioner's Approach, 6th edn. McGraw-Hill Science/Engineering/Math, New York (2004)
14. Rosemberg, C., et al.: Elicitação de Requisitos e Design Participativo através de Protótipos de baixa Fidelidade - um Estudo de Caso. In: Congresso Tecnológico InfoBrasil, Fortaleza, Brazil (2008)
15. Schwaber, K.: Scrum Guide. Scrum Alliance, Seattle (2009)
16. Snyder, C.: Paper Prototyping: The fast and easy way to design and redefine user interfaces. Elsevier, San Francisco (2003)
17. Sommerville, I.: Software Engineering, 6th edn. Addison-Weasley, Reading (2000)
18. Sy, D.: Adapting Usability Investigations for Agile User-Centered Design. Journal of Usability Studies, 112–132 (2007)
19. Vredenburg, K., et al.: A Survey of User-Centered Design Practice. In: Computer-Human Interaction 2002, Minnesota, USA (2002)

Towards a Conceptual Framework for Interaction Design for the Pragmatic Web

Heiko Hornung and M. Cecília C. Baranauskas

Instituto de Computação, Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, Cidade Universitária, 13.084-971, Campinas, SP, Brasil
{heiko,cecilia}@ic.unicamp.br

Abstract. In the current World Wide Web, useful information on web sites is often mixed with a lot of information that is not relevant to a user at a particular moment, or is presented in a format that is not optimal for a particular person using a specific artifact. In this paper we argue that to solve problems related to information relevance, presentation and flexibility of use, approaches are required that provide users with uniform ways of accessing and using information and services that are relevant to them at a particular moment in a way that suits their competences and needs. Informed by the Pragmatic Web and hence the questions of how and why people actually access information and services, this work proposes to set a basis for a conceptual framework to better understand, reason about, and design interaction in the Web.

Keywords: Interaction Design, Pragmatic Web.

1 Introduction

In the World Wide Web (Web in the following) useful information on web sites is often mixed with a lot of information that is not relevant to a user at that particular moment. Furthermore, information may be presented in a format that is not optimal for a particular person. Approaches to alleviate these problems include: designing web sites adhering to accessibility and usability guidelines, placing links to frequently accessed content prominently on a page, offering customization options, and offering a search function or a site map. These approaches are all site-specific and may be implemented differently or not at all in different sites, which means that a user has first to try to find her way around when entering a new site. This is a problem that affects all users, but especially those with less experience in Web use and the users of less frequently requested services (e.g. requesting a new passport after a loss).

User-specific strategies to alleviate the problem of retrieving relevant information include using browser bookmarks, memorizing URIs, or using site-external search engines. These strategies impose an additional cognitive load on the users, e.g. the organization of large bookmark collections or the localization of the desired result in the list presented by the search engine. Furthermore, in the case of search engines, users often have to phrase their queries using a vocabulary that is relatively similar to the one used on the desired site and potentially relatively different from their own.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 72–81, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Regarding the problem of information presentation, solution strategies often involve some form of end-user programming (e.g. user style sheets or web scraping [1]).

We argue that, in order to solve problems related to information relevance, presentation and flexibility of use, approaches are required that are independent of a specific web site or service provider insofar that mechanisms do not depend on a concrete implementation but provide users with uniform ways to access and use information and services that are relevant to them at a particular moment in a way that optimally suits their competences and needs.

A prerequisite for a solution is that a web site needs to provide means to analyze and process its contents computationally. Furthermore, we need to understand how people access and use information and services, as well as how they interact with other people or electronic services. Moreover, this understanding has to inform methods and techniques that can be utilized to effectively design those solutions.

Regarding the analysis and the processing of information in the Web, the HTML-based Web of today, which we will call “Syntactic Web” from now on, offers limited possibilities like analyzing the structure of a document. This makes it difficult to computationally process documents, because code would have to be adapted each time the source document structure changes. In contrast to the document-centric Syntactic Web, the Semantic Web [2] is centered on the meaning of and the relationship between data.

As to the challenge of understanding how people access and use information in the Web and how they interact with each other, the notion of the “Semantic Web” is required, but not sufficient: among the main concerns of the Semantic Web are data integration, interoperability, and automated electronic agents. To date, research that is concerned with user interaction in the Semantic Web often describes only visualization, navigation and search in semantically annotated data sets. The Pragmatic Web [3], on the other hand, permits to analyze the Semantic Web enabled interaction of people with each other or with services. The vision of the Pragmatic Web is “to augment human collaboration effectively by appropriate technologies”. Important topics are contextualized meaning, meaning negotiation, and the practices of virtual communities [4].

We hypothesize that, when adopting a Pragmatic Web perspective, the process of interaction design results in information and services that may be more relevant to people, that may use presentations that better fit people's needs, and that may provide a flexibility of use that accommodates a larger variety of competencies. In order to better understand, reason about, and design interactions in the Pragmatic Web, in this work we propose the basis of a conceptual framework. The paper is organized as follows: the next section presents our view of interaction in the Pragmatic Web, illustrated by a usage scenario, and points out how interaction in the Pragmatic Web differs from interaction in today's Web. After that, we present related work that might provide a theoretical and practical underpinning for such a conceptual framework. The subsequent section provides arguments why the definition of a conceptual framework in fact might contribute to the solution of the problems described in the introduction of this paper. The last section concludes.

2 Interaction in the Pragmatic Web

In order to illustrate how the perspective on interaction of people mediated by the Web shifts when introducing the notion of the Pragmatic Web, we describe a short scenario of interaction in the Pragmatic Web and provide a discussion of the differences to interaction in today's Web. While Syntactic Web scenarios are focused on users interacting with other users or with digital artifacts, Semantic Web scenarios often include electronic agents that assist users or execute tasks on behalf of users interacting with other electronic agents or users. As the Semantic Web often is seen as an enhancement and not a replacement of the Syntactic Web, our vision of the Pragmatic Web is that of a Web that builds on the Syntactic and Semantic Web, i.e. uses the respective stacks of protocols, methods and tools. Consequently, Pragmatic Web scenarios include users and electronic agents, but introduce aspects like relevance, intention, or negotiation.

2.1 A Scenario of Interaction in the Pragmatic Web

Alice, an elementary school teacher, still remembers the time when she had to maintain various accounts at different social network, photo sharing or messaging services in order to keep in touch with her friends. Today, when she wants to send a short message to one of her friends, she does not have to worry which social network or messenger he is using. When she takes a photo with her camera or browses her own photo gallery, she can share a photo without having to switch to the client of the photo sharing service or enter their site. She does not have to worry that the principal sees potentially embarrassing pictures of her last birthday party because the fellow teachers that are within her circle of friends are aware that those pictures are not meant to be distributed at work.

Last week, Alice received a reminder from her local government agency, informing that her passport is soon expiring. The reminder contained a list of necessary documents together with the new specifications for the passport photograph. When browsing her photo gallery, Alice is now able to automatically verify if a photo is a valid passport photograph.

Later in the afternoon, she has an online meeting, where she and other teachers discuss the learning process of students, exchange didactic material, review test results, etc. Until some time ago, Alice was forced to use the commonly agreed upon Content Management System to upload files or discuss cases in a forum. Now, she is able to use the same tools that she uses for e.g. sharing photos and videos or chatting with her friends. The files produced during the online meetings or other activities are organized in a manner that suits Alice's needs, while other teachers organize them according to their needs. Although Alice is using some tools she also uses for private purposes, she is always aware of the work context when engaged in activities with other teachers.

2.2 Discussion of the Scenario

The base concepts in the scenario presented above are people acting as users of services and creating or manipulating content. In today's Web, the scenario described above would not be possible for various reasons. Users are identified by accounts, i.e. they have to maintain and remember information about different user names, passwords, password recovery questions, etc. Services provide operations and can be accessed via different user interfaces. The content created by a user using a service usually stays within the realm of the service provider (of course, the content can be copied and uploaded to the space of a different service provider). The service provider usually defines the user interfaces and the representation of the content (e.g. if you upload a video to YouTube, the video will be converted to different formats with different resolutions that can be displayed by the player at youtube.com). This means, that Bob, a friend of Alice's, would not be able to view her photo using another service than the one Alice used to upload her photo. In order to view her photo, either Bob has to have an account with Alice's service provider, or Alice's service provider has to accept a single-sign-on solution like OpenID (http://openid.net) and Bob has to be a user of this solution. Alice in turn may be able to access her photo via a different service, if all involved services support OAuth (http://oauth.net) or a similar protocol. If Alice would like to share a photo with Bob within the context of a working group at school, i.e. would not like him to share the photo with other friends, she would have to add a comment to the photo or notify Bob in a separate message.

In order to put the scenario in practice, syntactic or semantic approaches (e.g. providing access to services via APIs or enabling data portability or interoperability by providing semantic data descriptions), although necessary, would not suffice. Additionally, methods and techniques are required that allow to put meanings into context and enable meaning negotiation and the analysis and design of practices of virtual communities.

3 Related Work

We see the “Pragmatic Web” in the intersection of the three major areas of Human-Computer Interaction (HCI), Information Systems (IS), and Web Science. From our own perspective of interaction design, different topics from each area will influence and inform the proposed conceptual framework. In this section we will first explore contributions from HCI, IS and Web Science that are candidates for the theoretical foundation of the conceptual framework.

HCI research has acknowledged a long time ago that humans are not simply components of a system that can be studied in isolation in a laboratory environment [5]. This has led to what some researchers call “post-cognitivist theories”, theories that go beyond the study of cognitive abilities, and that have a substantial amount of quantitative and significantly less qualitative elements than more traditional HCI theories rooted in cognitive psychology, human factors or ergonomics. Post-cognitive theories and models often start from the notion of language and the notion of action as a form of language use. Examples are activity theory, distributed cognition, actor-network theory, phenomenology (see [6] for a comparison of the four theories from the point of view of activity theory), or the language/action perspective [7]. Some of these theories are also employed in the field of information systems research [8].

“Information Systems are a multi-disciplinary subject, whose objects of study are information and its functions, information technology and its use in organizational contexts" ([9], page xi). Understanding "organizational contexts" not only as relationships between people in the formal work context, but as relationships between people interacting together towards some end, theories and methods from the field of information systems research that are concerned with the use of IT artifacts are relevant to our proposal. One valuable contribution to our proposal is the fact that in information systems research social aspects that go beyond the direct interaction of people with IT artifacts have always been a concern. One of the theories and frameworks that consider the use of IT artifacts from a technical as well as a social perspective is Organizational Semiotics [10]. It focuses on understanding the different properties of signs on various levels to analyze and design information systems in terms of three human information functions: expressing meanings, communicating intentions and creating knowledge.

Besides Organizational Semiotics, other approaches exist that consider how humans use IT artifacts in organizational contexts, and different efforts to compare, integrate or synthesize those approaches have been made (e.g. [11]). One notable example is socio-instrumental pragmatism [12], a proposal of an action oriented theory for IS research that synthesizes different action theories and thus is better able to cover different demands from IS research than a single theory. [12] further describes different action themes and their theoretic roots, i.e. purposeful, social, interactive, mediated, creative or situational action.

Web Science [13] is an emerging interdisciplinary field that sets out to understand the Web not only as a delivery vehicle for content, but as an object of study in its own right. Besides technical or engineering issues like the Semantic Web or web services, also the social aspects like Web use or governance are considered. Among Web Science research questions that are also relevant to our proposal are those about significance, relevance, reputation and trust.

The Semantic Web has been proposed as an extension to the current Web with the intent to introduce meaning to Web pages, processable by human or machine agents [2]. However, the augmented semantic contents to a great extent remain inaccessible or unintelligible for human agents. Some authors claim that considering meanings is necessary, but not sufficient, and that the purpose and context of information also has to be considered (e.g. [14]). Singh thus states that the vision of the Semantic Web can be implemented via Pragmatics, a branch of Semiotics that deals with context-based meaning [15].

Building on the Semantic Web, the Pragmatic Web sets forth "to augment human collaboration effectively by appropriate technologies" [3]. Although there does not yet exist a commonly accepted definition for the term, research topics comprise contextualized meanings, meaning negotiation, and the practices of virtual communities. Hornung and Baranauskas describe the significance of the Pragmatic Web for interaction design [16].

An important concept of Pragmatic Web research is related to intentions that lead to meaningful actions. Within cognitivist HCI theory, translating intentions into actions in order to realize goals using a tool is part of bridging the "Gulf of Execution" [17]. Among the post-cognitivist HCI theories that conceptualize humans as subjects acting intentionally mediated by tools are activity theory and phenomenology [6]. Regarding the analysis and the design of actual actions, the concepts of usability, affordances, and actability exist, whereas the term "affordance" has different significations in the HCI and IS communities [18][19]. [16] comment on the different notions and their relevance for interaction design.

After having given a brief overview of general related work in the three areas that influence and delimit a conceptual framework for interaction design in the Pragmatic Web, we will point out more specific topics that provide complementary approaches or pointers to answers for the question of how to implement solutions informed by a conceptual framework for interaction design in the Pragmatic Web.

Semantic Web User Interaction is the name of a community that tries to foster the dialogue between the Semantic Web and the HCI communities. Recurring topics in literature are navigation and search in and visualization of structured datasets, semantic annotation of web pages, creation of semantic data, as well as contextualizing and customizing user interaction [20]. Literature about the Social Semantic Web explores the question of how to apply Semantic Web technologies to the Social Web, combining “the best of the two worlds”, i.e. facilitating navigation and searching by semantically annotating content on the one hand and promoting the creation of structured content by using Social Web mechanisms on the other. Main topics in literature are questions of interoperability and integration [21][22]. The subject of both Social Semantic Web and Semantic Web User Interaction literature is the interaction of humans in a semantically enriched Web, whereas Social Semantic Web literature focuses on questions like interoperability and integration, while Semantic Web User Interaction literature focuses on interaction design. Both strands have the potential to bridge the gap between the often technically and computationally oriented Semantic Web community and the, at this stage, rather conceptually oriented Pragmatic Web community.

4 The Pragmatic Web as Proposed

In the scenario of Section 2, people have multifaceted identities and exhibit facets of their identities to other people. Depending on the context, two facets of a single person might even contradict and thus give the impression of different identities. Services enable different operations for different user interfaces. They are independent of content, which implies that users have more freedom to choose the user interface with which they access content and the content presentation that is most adequate. Terms of service and terms of use determine rights and obligations of service and content providers and consumers. A content item has an author and different presentations. The context determines under which conditions people may access content in which way. Customization determines which services are used to access which content presentations in what way. The division into the five dimensions people, services, content, customization and context is depicted in Figure 1, which shows two fictitious services, each with a different set of operations, user interfaces and terms of service. Depending on the facet of the accessing person's identity, the terms of service, the terms of content use, the service user interface and chosen operation, not all presentations of a content item might be available.

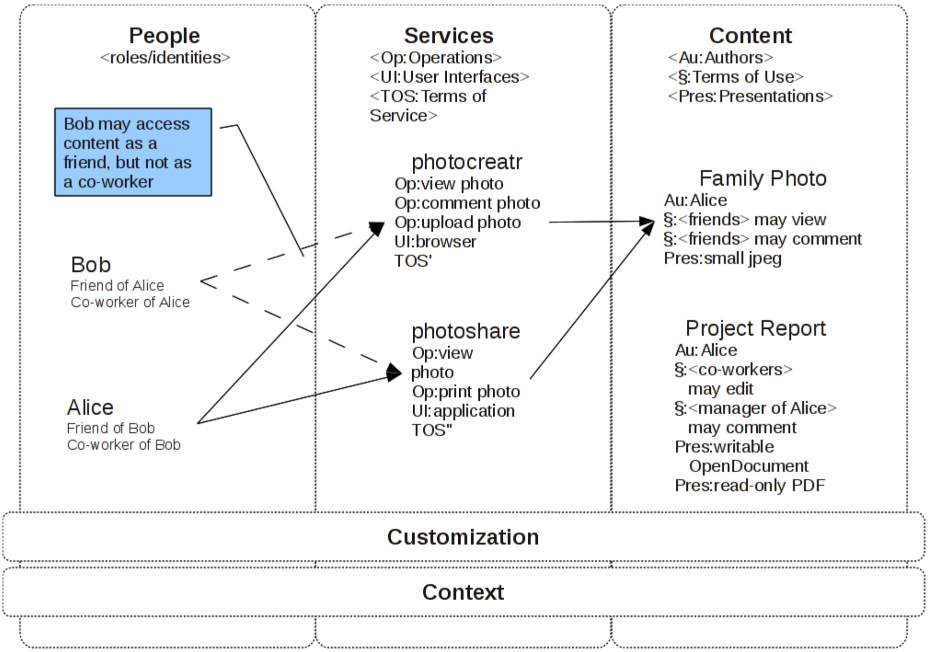


Fig. 1. *People* access *content* independent of *services* but sensitive to *context*

The usage of digital artifacts is not necessarily a primary activity during work and users are not necessarily experts regarding the use of the respective digital artifacts. This point is important regarding our further approach. Breakdowns during artifact interaction of expert users often occur because of usability or accessibility problems of a specific artifact. When considering the interaction of users with a variety of artifacts, e.g. different web sites created by different authors, breakdowns can occur because of what a usability analysis of a single artifact would label “lack of consistency” [23]. Only in this case, each artifact could pass a usability evaluation individually and even if the same or similar design principles had been used during the creation of each artifact, inconsistencies are likely to arise when considering a set of artifacts authored by different entities.

When introducing the Pragmatic Web, the focus hence moves from isolated web pages in the Syntactic Web and semantically structured data in the Semantic Web to the question of how and why people actually access information and services. For example, while teachers create or upload video, audio, text or other documents in the Syntactic Web and semantically annotate those files in the Semantic Web, they now share and access materials such as interviews, test results, educational games, etc.

We believe that this shift of perspective enables us to create artifacts that are more relevant and meaningful to a wider range of people. One might argue that, taking a user-centered or participatory design approach would already result in relevant and meaningful artifacts. This might be true for artifacts that are used by a limited or homogeneous set of people. However, even if many users achieve an acceptable performance, often the created artifact is not optimal for most of them [24]. For example, web sites of tax authorities have to accommodate for the whole set of requirements of a country's citizens, corporations and other entities, while usually only a small fraction of the whole site is relevant for a single citizen [25].

In the Pragmatic Web, the interaction of people mediated by digital artifacts is substantially different from the interaction in the Syntactic or Semantic Web. In the Syntactic Web, interaction is usually based on web sites (in the case of browser based interaction) or on service provider based interaction (in the case of e.g. instant messaging services). In the Semantic Web, interaction is based on datasets. In our vision of the Pragmatic Web, interaction is based on the intentions of people which are materialized by actions (we do not yet want to enter the discussion whether to call the concept action, activity, act, etc.), i.e. interaction abstracts from a service provider who enables the actual action and gives people the control on how the results of their actions are presented and with whom and under which conditions they are shared.

In order to understand and design interactions in the Pragmatic Web, we aim at defining a conceptual framework, the construction of which will be informed by different theories and models from the following areas: HCI, because we need to analyze and design the interaction of people with digital artifacts or with each other mediated by digital artifacts; IS, because we need to understand how people access and process information in an organizational context, whereas by organization we mean any relationship between two or more people; Web Science, because we need to understand the basic mechanisms and the still many open questions of the Web. The base elements of the conceptual framework will comprise people, services, and content. Context and customization are considered as orthogonal dimensions. Among other important concepts that might have to be considered in the conceptual framework are values like trust, privacy, and authority.

6 Conclusion

In this work we addressed the problems of information relevance and presentation as well as flexibility of access and use of services and information. We hypothesized that a Pragmatic Web perspective can contribute to the solution of these problems by providing means for understanding how people access and use information and services and how they interact with each other in the Web. We presented our vision of interaction in the Pragmatic Web and proposed to develop a conceptual framework for interaction design for the Pragmatic Web informed by HCI, IS, and Web Science. The framework enables the design of interactions, in which people can collaborate while each participant is accessing services and content presentations that best fit his/her needs.

Considering the implementation of solutions that follow the proposed framework, a number of challenges arises regarding the five dimensions: people, services, content, customization, and context. Those challenges range from technical (e.g. protocols or standards) to formal organizational (e.g. forms of meaning negotiation) and informal pragmatic and social challenges (materialize intentions into actions, awareness of effects of an action) and include questions related to interaction design. One of our next steps is to identify and analyze those challenges.

Future work includes a case study where the theoretical conceptualizations can be materialized into a proof-of-concept implementation. In order to be feasible, this proof-of-concept implementation will only implement a small fraction of the framework. The actual part of the framework that will be implemented and the concrete implementation depend on the requirements of a research project in which the authors are currently involved.

Acknowledgments. This work is funded by the Brazilian National Council of Scientific and Technological Development (CNPq, grants 141058/2010-2 and 560044/2010-0) and by the Coordenadoria de Aperfeicoamento de Pessoal de Nivel Superior (CAPES, grant 01-P-08503/2008). The authors also thank their colleagues from InterHAD for insightful discussions.

References

1. Huynh, D., Mazzocchi, S., Karger, D.: Piggy Bank: Experience the Semantic Web Inside Your Web Browser. Web Semant. 5, 16–27 (2007)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284, 34–43 (2001)
3. Schoop, M., de Moor, A., Dietz, J.L.G.: The Pragmatic Web: a Manifesto. Commun. ACM 49, 75–76 (2006)
4. Pohjola, P.: The pragmatic web: Some key issues. In: Proceedings of the 6th International Conference on Semantic Systems, I-SEMANTICS 2010, pp. 30:1–30:8. ACM, New York (2010)
5. Bannon, L.: From Human Factors to Human Actors: the Role of Psychology and Human-Computer Interaction Studies in System Design. In: Greenbaum, J., Kyng, M. (eds.) Design at Work: Cooperative Design of Computer Systems, pp. 25–44. Lawrence Erlbaum Associates, Hillsdale (1991)
6. Kaptelinin, V., Nardi, B.A.: Acting with Technology: Activity Theory and Interaction Design (Acting with Technology). The MIT Press, Cambridge (2006)
7. Schoop, M.: An Introduction to the Language-Action Perspective. SIGGROUP Bull. 22, 3–8 (2001)
8. Goldkuhl, G.: Socio-instrumental Pragmatism: A Theoretical Synthesis for Pragmatic Conceptualisation in Information Systems. In: The 3rd Intl Conf on Action in Language, Organisations and Information Systems ALOIS, vol. 3, pp. 148–165. Limerick University, Limerick (2005)
9. Liu, K.: Semiotics in Information Systems Engineering. Cambridge University Press, New York (2000)
10. Stamper, R.K.: Organisational Semiotics: Informatics Without the Computer? In: Liu, K., Clarke, R.J., Andersen, P.B., Stamper, R. (eds.) Information, Organisation and Technology: Studies in Organisational Semiotics, pp. 115–171. Kluwer, Dordrecht (2001)
11. Cordeiro, J., Filipe, J.: Language Action Perspective, Organizational Semiotics and the Theory of Organized Activity – a Comparison. In: Proceedings of the Workshop DEMO, Tilburg, the Netherlands (2003)
12. Goldkuhl, G.: Socio-instrumental Pragmatism: A Theoretical Synthesis for Pragmatic Conceptualisation in Information Systems. In: The 3rd Intl Conf on Action in Language, Organisations and Information Systems ALOIS, vol. 3, pp. 148–165. Limerick University, Limerick (2005)

13. Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N., Weitzner, D.J.: A Framework for Web Science. Found. Trends Web Sci. 1, 1–130 (2006)
14. de Moor, A., Keeler, M., Richmond, G.: Towards a Pragmatic Web. In: Priss, U., Corbett, D.R., Angelova, G. (eds.) ICCS 2002. LNCS (LNAI), vol. 2393, pp. 235–249. Springer, Heidelberg (2002)
15. Singh, M.P.: The Pragmatic Web. IEEE Internet Computing 6, 4–5 (2002)
16. Hornung, H., Baranauskas, M.C.C.: An Interaction Design Perspective on the Pragmatic Web: Preliminary Thoughts. In: Paschke, A., Weigand, H., Behrendt, W., Tochtermann, K., Pellegrini, T. (eds.) Proceedings of the 5th Intl. Conf. on Semantic Systems (I-SEMANTICS 2009), Graz, Austria, September 2-4, pp. 695–705 (2009)
17. Norman, D.A.: Cognitive Engineering. In: Norman, D.A., Draper, S.W. (eds.) User Centered System Design, pp. 31–61. Lawrence Erlbaum Associates, Hillsdale (1986)
18. Ågerfalk, P.J., Cronholm, S.: Usability versus Actability: A Conceptual Comparative Analysis. Poster Sessions: Abridged Proceedings of HCI International, New Orleans, LA, USA, pp. 235–237 (2001)
19. Goldkuhl, G.: Actability Theory Meets Affordance Theory: Clarifying HCI in IT Usage Situations. In: Golden, W., Acton, T., Conboy, K., Van der Heijden, H., Tuunainen, V.K. (eds.) 12th European Conference on Information Systems, Galway, Ireland, vol. 16, pp. 790–801. AIS (2008)
20. André, P., schraefel, m., Van Kleek, M., Karger, D.: Bringing the Semantic Web Home: a Research Agenda for Local, Personalized SWUI. In: CHI 2008, SWUI Workshop (2008)
21. Breslin, J.G., Decker, S.: The Future of Social Networks on the Internet: The Need for Semantics. IEEE Internet Computing 11, 86–90 (2007)
22. Bojārs, U., Passant, A., Breslin, J.G., Decker, S.: Social Network and Data Portability Using Semantic Web Technologies. In: Flejter, D., Grzonkowski, S., Kaczmarek, T., Kowalkiewicz, M., Nagle, T., Parkes, J. (eds.) BIS (Workshops). CEUR Workshop Proceedings, vol. 333, pp. 5–19. CEUR-WS.Org (2008)
23. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1993)
24. Dix, A.: Human-Computer Interaction: A Stable Discipline, a Nascent Science, and the Growth of the Long Tail. Interacting with Computers 22, 13–27 (2010), Special Issue: Festschrift for John Long
25. Bonacin, R., Baranauskas, M.C.C., Liu, K., Sun, L.: Norms-based simulation for personalized service provision. Semiotica 2009, 403–428 (2009)

Blowing Light: Green-Based Interaction Design

Yu-Chun Annester Huang*, Chih-Chieh Tsai, Teng-Wen Chang,
Pen-Yan Tsai, Tien-Hsin Hung, and Jai-Jung Chen

Dept. of Digital Media Design, Natinal Yunlin University of Science and Technology,
No. 123 University Road, Section 3, Douliou, Yunlin, Taiwan
{AnnesterHuang,thedogtsai}@gmail.com,
{tengwen,u9635001,u9635010,u96350324}@yuntech.edu.tw

Abstract. As green being a significant issue of these years, we want to discuss about how to combine green with technology through ambient design. Not only use intelligent methods but also interact with fun playing interaction. The relationship between people and the feedback of installation can make people help themselves to meet the needs of them. This paper provides an innovation concept of ambient Intelligence. People can help themselves through a wind-blower installation by conventional power-generating method.

Keywords: Ambient Intelligent, Green, Human Power, Wind-Blower, Light, Interactive installation.

1 Introduction

Human interaction is often starting by studying how human wants in a specific situation[1]. With technologies embedded in everyday's objects, while more such objects are developed, the smarter materials and environment has become a reality around us. Ambient interaction in current computing approach creates a new paradigm in design as well as computing. On the other hand, with the global energy problems, the Green is an immediate issue that we need to address in our environment and that is often conflicted with the normal energy consumption nature of computing.

The problem of combination between Green and technology (especially computing) becomes more and more significant. With the Ambient Intelligence, designers and scientists explore a vision of future daily life that can apply the Green and technology together. What we want to develop is how to make people help themselves with an installation that embedded to the environment (stair in the prototype in this research) by a conventional making fire gesture.

In the beginning of this prototype, the concept was based on studying the human power activities for gaining energy that we found is making fire. Wind-blower or hand drilling are studied for understanding the gestural details of making fire. With human power, the activity of making fire will remind users the importance of energy and the consumption behind the lighting. Further, the interesting concept and tool design of this installation create a fun interaction that will achieve the concept of reuse, meaningful, and high feedback for the users.

* Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 82–91, 2011.
© Springer-Verlag Berlin Heidelberg 2011

1.1 Ambient Intelligence as Interaction Theory

In an ambient intelligence world, the everyday's devices will support people in carrying out their daily routines, tasks and rituals in easy [2]. The natural reaction for using information and intelligence hidden in the network are the key to connect these devices. As these devices become smaller, connected and more integrated into our environment, the technology embedded in the devices and the surroundings will become transparent to the users[3]. In addition, ambient intelligence provides electronic environments that are sensitive and responsive to the presence of people. People are presented with the tools and the processes that are necessary in order to achieve relaxing interactions with this environment[4]. Therefore, ambient intelligence is suitable for defining the computing framework for building up our prototype, thus our interaction theory.

1.2 Green Power as the Media

Green Power is the power using non-energy-consumption source to generate energy [9]. Human power is often used as the primary source. As theory of a new media interface, human power should not only provide the energy needed for the installation but also interactive experience that comes along with personal evolvement. We want to discuss how to use green power as the media to implement the interaction theory we used, ambient intelligence.

Logically speaking, green power has to be adapted as the major purpose of this project. Thus people should work this installation through their bodied exercise, they need to help themselves to meet what they need for reducing the waste of energy.

1.3 The Problem of Our Interaction Design

The problem of combining green with technology lies on the familiarity of relating human power with interactivity. How users experience with human power and relate it to the interaction will affect the interactive experience of human in our prototype. We will use several methods to overcome these problem and learning the bits and bells during the implementing our prototype.

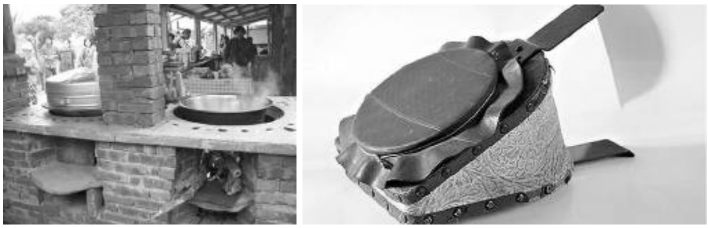
2 The Concept

2.1 Motivation

The major motivation for this prototype is how ambient interaction design can be integrated and triggered by human power as an experience. Interaction design cannot be fully realized and understood without the prototypes produced. Therefore, for testing the green-based interaction design with human power-triggered, we developing a prototype called *Blowing Light*.

2.2 Concept

The Concept of our interaction design is to make people use a conventional power-generating method, blowing for fire to turn on the light (As Fig. 1) is found after testing several alternatives. Additionally, this design concept invokes a mode of interaction that make people aware a well-known saying, “No pay, No gain”. By doing so, people or our users will be more aware of Green issues implicitly in our interaction design.



(a) (b)

Fig. 1. (a) Traditional method of blowing for fire (b) wind-blower

2.3 How We Approach

For implementing the prototype described above, we use three steps: (1) reviewing/analyzing relevant technology and cases, (2) exploring the interaction activities and (3) implementing with a prototype: Blowing Light that are described as followed.

1. At first, we analyze the design researches for green and study some relative projects about saving energy. Since there are only few that are implemented, most of relevant studies lie upon the cases they develop. Therefore, the case studies are the key method used in this step. By exploring those cases, we discovered they use human power and tools to reduce the waste of natural power like electric that achieve the combination of green and technology that was the problem we want to solve.
2. Secondly, the basic model of design strategy is formed that was based on the principles we gained from the analysis above: such like green, playing and human power. Besides, we investigate the intuitive reaction with testing from university level students.
3. Finally we developed a special interactive method for our users (in this case, university students) to turn on the light at night. This installation was built up with some green materials like bamboo wind blower and recycle plastic cup as light lampshade, reuse glass papers for decorating the lampshade.

3 Case Studies

Three cases are selected based on Green, playing and human power. They are (1) *the Fun Theory—Piano staircase,* (2) *Water for people—Playpumps*, and (3) *Playmade energy—energee-Saw*. Each will be described as followed:

3.1 Case 1: The Fun Theory—Piano Staircase[5]

The funtheory.com of the Volkswagen has created a platform to attract a lot of interesting projects with fun as their center motivation that is close to what we want to develop in Blowing Light. Piano Staircase video on *Youtube* has been the famous ambient interaction design since two years ago; it really makes people want to take the stairs instead of taking escalator through a fun refit piano staircase. (As Fig. 2) This project not only makes people take action but also have a lot of fun.

(a) (b)

Fig. 2. (a) The fun theory installed the piano staircase (b) People took the piano staircase instead of escalator. (All images are courtesy of The Fun Theory `http://www.thefuntheory.com/piano-staircase`).

3.2 Case 2: Water for People—PlayPumps [6]

PlayPumps [6] is an interesting project implemented by Water For People. This simple system has the power to change whole communities, from health and education, to employment and economy. The same basic design can be adapted for all manner of situations. It's all thanks to the power of children at play! The design of the PlayPumps (As Fig. 3) water system makes it highly effective, easy to operate and very economical, keeping costs and maintenance to an absolute minimum.

As seen in Figure 3, while children have fun spinning on the PlayPump merry-go-round (1) clean water is pumped (2) from underground (3) into a 2,500-liter tank (4)

(a) (b) (c)

Fig. 3. (a), (b) Children was playing the PlayPumps for providing the power to (c) push water into the tank. (All images are courtesy of PlayPumps International: `http://www.waterforpeople.org`).

standing seven meters above the ground. A tap (5) allows access to the stored water. Any excess water pumped up is diverted back down the borehole (6). The storage tank (7) provides a rare opportunity in rural villages to advertise – both as a commercial billboard, and also for health and educational messages. Revenue from the billboards is used to pay for pump maintenance. The system is capable of drawing 1,400 liters of water per hour at 16rpm from a depth of 40 meters.

3.3 Case 3: PlayMade Energy: The Energee-Saw[7]

The Energee-Saw from PlayMade Energy (As Fig. 4) used the local materials to build the main structure of the see-saw. This greatly reduces logistical costs and the carbon footprint of transportation. It also engages the local community into building the product, fostering a pride of ownership and providing work. Energee-Saw will power low-drain electrical appliances such as LED based classroom lighting, radios and mp3 devices, mobile phones and potentially low-wattage laptop computers.

(a) (b) (c)

Fig. 4. (a) Designer Daniel Sheridan (b)(c) Children was playing the Energee-Saw for providing the power.(All images are courtesy of PlayMade Energy. `http://www.playmadeenergy.com/`)

3.4 Analysis

Basic on the cases upon, we can see that they redesign the interface and make people easier to understand how to use the installation through their experiences. The comparison of each case in terms of input/output/behavior/purpose/material is shown in Table 1. There are some principles gained from the case studies are shown as followed:

Table 1. Interaction Analysis list of related cases above

Interaction analysis			
Case	**Piano staircase**	**PlayPumps**	**Energee-Sew**
Input	Human step stress	Human pushing power	Human power
Output	key tone	Push water to tank	Provide electric
Behavior	Playing piano stairs	Playing merry-go-round	Playing Energee-Sew
Purpose	Make people walk the stair instead of escalator	Provide electric	Provide electricity
Material	Design materials	Design materials	Native recycle or reuse materials

(1) Green: Conventional objects have physical characteristics; mechanical ones also have capabilities[8] so we tried to indicate the green to traditional but good method. However indicate technology to some techniques that need to build up with green materials such like LED light. (As Fig. 5)

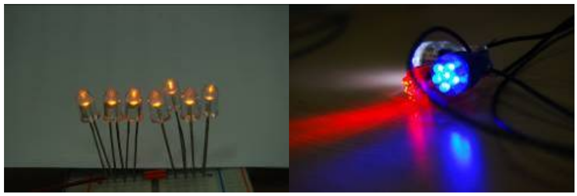
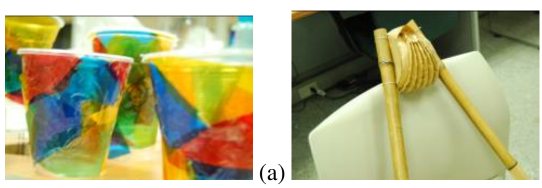


Fig. 5. LED light we used as green materials

(2) Playing with Human power: We re-defined a behavior that was burning the woods for lighting darkness, however, the waste of wood was not that green for the earth now, therefore we kept the behavior that using wind-blower to blow the fire prior of time to be the interaction.(Table 1)

(3) Recycle or reuse material: Those lights were made of recycle plastic cups which were used plenty in Taiwan, the skin of cups were decorated by reuse colorful glass papers for achieving the goal of combine green with technology. (As Fig. 6)



(a) (b)

Fig. 6. (a) The skin of lights is made of reuse glass papers. (b) The wind-blower used green materials, bamboo as materials.

4 System Design

People playing with generated power like using air-blower for providing electronics to make the light open that making a zero-energy high-tech playground. The behaviors analysis has conducted for finding the details and intuitiveness of gesture we need.

4.1 Behavior Analysis (The Interactive Process)

When people need lights, they will try to find out some light source or tools for lighting up the dark place. When they find out some light source, they will try to keep it lighting or make it stronger for lighting all of the space around. Base on this point, we used a LED light as the light source in the dark and drew user's attention to find out the wind-blower, if they took up the blower they would try to blow for keeping the light source, LED light. After blowing, all of lights would open for the user and let user pass safety.

4.2 The Location

We choose a dark and narrow place of school's environment for testing the lighting effects (As Fig. 7) and creating a scenario for motivating people. The installation is set up on the stairs of college of design NYUST where it is very dark during the night and a lot of students stay late for working in the labs or office need to go passed this staircase. Consequently, most of students would take elevators instead. The purpose of Blowing Light not only concerns about the security of students but also makes students get used to take the stairs.

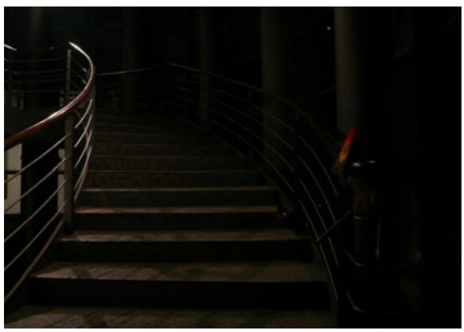


Fig. 7. School's environment where is testing the lighting effects

4.3 The System Process

Blowing light is comprised of three parts (a) user transforms human power into the input signals to sensor, the transformation process is through wind from wind-blower, (b) system computes collected signals and, (c) makes a feedback at LED lights. The light timing of LED is according to the time interval of channel circuit. This concept is that sensor detects the turning speed of fan. When the speed is over the value we set, control component will make the LED light start to shine. (Fig. 7)

4.4 Interactive Process

User passes by the installation and takes the wind-blower to blow the sensor, the controller will accept computed signal from user's energy. The time user blows will determine how long LED lights will light for user and make user pass safety.

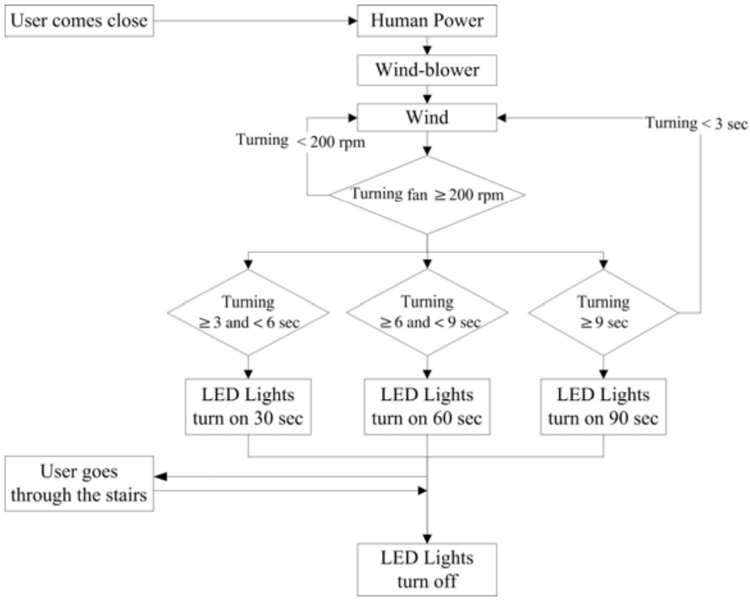


Fig. 8. For turning LED lights on, we used circuit sensor to predicate when user will need and how long LED lights will need to keep on

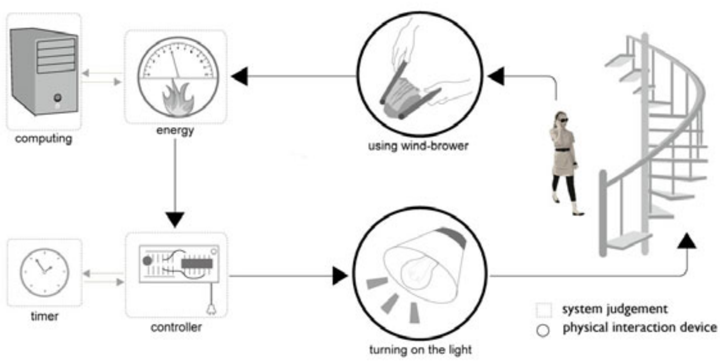


Fig. 9. User provides power to installation by playing the wind-blower

5 Scenarios

For understanding the usages of Blow-lighting system, one scenario is described as followed: blowing the darkness of path away.

Molly is a 20 year-old girl, every time she goes back from school, she needs to pass through a narrow and dark stairs which make her scared on the way home. One day, she was going back home and walking to the stairs as usual, she was trying to run over but she fell down on the floor. Suddenly she found out a wind-blower and many

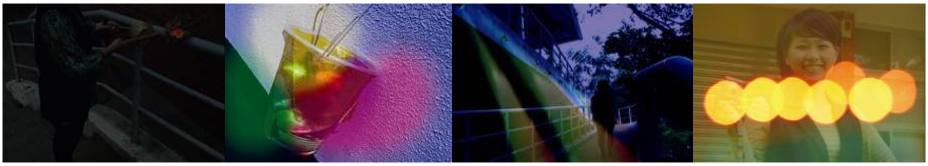


Fig. 10. User Molly was playing Blowing Light installation

maple leaves which as red as fire shining on the handle of stairs, there were little weak lights shining behind the maple leaves so she took up the wind-blower and started to push. (As Fig. 9)

Molly wanted to know if the light will be brighter after she was pushing stronger. When she tried her best to push the wind-blower, she thought that it was so funny to do this, but the colorful lights (As Fig. 9) suddenly bright at that moment, everything was so clear. Molly felt so happy that she didn't need to worry about her security around this stairs anymore. Finally she got home with appreciate. (As Fig. 9)

6 Conclusion

We used conventional method of blowing light as initial concept and combined green and technology through ambient design. In the process of making prototype, we used recycle materials for decorating the skin of lights for achieved another concept, Green. In other words, we used less cost to procure a zero-energy high-tech playground.

On the other hand, interactive interface design was based on most intuitive and natural thinking of user. Furthermore, we used fun interaction for user enjoying the time of playing installation even it cost human power to work on. Our prototype is not only artistic but also practical and provides a safety solution for user.

In the future, we plan to research other movement of body as human power which provides the energy of installation.

References

1. Lee, J.: Interactive media- A multimodal Approach to interface Design for Human- Computer Interaction in Digital Design Environment (2002)
2. Punie, Y.: A social and technological view of Ambient Intellgence in Everyday Life: What bends the trend? (EMTEL Network Key Deliverable). Institute for Prospective Technological Studies, Directorate General Joint Research Centre, European Commission (2000-2003)
3. Winkler, T.: Composing Interactive Music: techniques and ideas using Max. Massachusetts Institute of Technology (1998)
4. Kameas, A., Mavrommati, I., Markopoulos, P.: Computing in tangible: using artifacts as components of ambient intelligence environments. In: Riva, G., Davide, F., Alcaniz, M. (eds.) Ambient Intelligence: The Evolution of Technology, Communication and Cognition, pp. 121–142. IOS Press, Amsterdam (2004)
5. The Fun Theory, `http://www.thefuntheory.com/piano-staircase`

6. Obiols, A.L., Erpf, K.: Evaluation of the PlayPumps. Rural Water Supply Network (RWSN) (2008)
7. PlayMade Energy, http://www.playmadeenergy.com
8. Snyder, J.: Energy and equity and human power: Ivan Illich Reviewed. Human Power 13, 10–11 (1998)
9. Schafer, R.: Scientific bases of human-machine communication by voice. Human-Machine Communication by Voice 92, 9914–9920 (1995)

The Impact of Rich Application Frameworks on User Experience Design

Tobias Komischke

Infragistics, Inc., 2 Commerce Drive, Cranbury, NJ 08512, USA
tkomischke@infragistics.com

Abstract. Rich internet and desktop application frameworks provide advanced features like direct object manipulation and advanced animations that can enhance the user experience. This paper traces the history and evolution of rich applications and in front of the background of the T.O.T.E. framework discusses the benefits that rich application capabilities provide for the human-system interaction.

Keywords: User Experience, Design, Rich Interactivity.

1 Introduction

When the first graphical user interfaces (GUIs) entered the market in the 1980s the use of pictorial and metaphoric user interface (UI) elements were revolutionary and significantly contributed to higher computer adoption rates. Direct manipulation of UI elements was realized through clicking and dragging and was performed on basically the same set of interactive UI elements that we still have today, such as push buttons and scrollbars. The behavior of these elements was mainly binary: on/off, show/hide, open/close.

Today, the input methods (clicking, dragging) are still the same for mouse-based applications (obviously, we now have multi-touch capabilities on more and more systems), yet the output behavior that the user interface is capable of showing is much richer. Features like drag & drop, fading, sliding, and transitions allow for more nuanced system output. This can add to the user experience both from a pragmatic and a hedonic standpoint and will be explained in the sections below.

It's worth looking at the history of rich interactivity as it shows that two areas of software have been influencing and boosting each other: the desktop world and the browser world. Before web browsers were invented in the 1990s, personal computing has been done based on desktop applications such as the Microsoft Office suite. The first browser-based user interfaces were static web pages with hardly any option to directly manipulate UI elements other than entering ACII onto forms. Later Adobe Flash made it possible to provide multimedia content that could also be interacted with. It took about 15 years until the first browser-based applications were available that could match the interaction capabilities of desktops. These so-called Rich Internet Applications (RIAs) were realized through technologies such as Flash and Java and were driven by the need of optimizing the appeal of web-based content presentation (multimedia integration, sophisticated animations, etc.). The typical websites that

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 92–97, 2011.
© Springer-Verlag Berlin Heidelberg 2011

pushed this development forward were marketing and advertising sites. Since then, more technologies like Ajax, Adobe Air, Microsoft Silverlight were brought to the market and allowed an ever growing number of designers and developers to create RIAs whose interactivity went beyond the desktop world.

With RIAs more and more becoming the gold standard for the way user interfaces should look and feel, the desktop world in recent years adopted the concept of rich interactivity under the name of rich desktop application (RDAs).

As of today, desktop and web applications have converged and now are hardly distinguishable. Development platforms like Microsoft Silverlight allow running web applications outside of a browser like a desktop application. Because of this, for the rest of this paper the term “rich applications” is used instead of distinguishing between RIAs and RDAs.

There is a fine line between boosting and wrecking the user experience of rich applications. Rich development platforms together with 3rd party controls enable programming highly sophisticated interactive features with a minimum of effort. Yet, even if something is easy to program and rich in its presentation, is it really useful, usable and desirable?

2 Benefits of Rich Interactivity

2.1 Cognitive Psychology Background

Human-system interaction can be considered as a sequence of problem-solving tasks. One of the earliest and most fundamental models is T.O.T.E. which stands for “Test – Operate – Test – Exit” (Miller, Galanter, Pribram, 1986). As a first step an assessment (test) about the current system state is made. If the system is not in the targeted state already, an operation (e.g. a user input) is executed to change the current state towards the desired state. Another test is made to see if the targeted state has been reached or not. More recurrences are performed until the test result is positive. Only then the sequence ends.

Each T.O.T.E. unit can be decomposed into sub-processes:

- Test:
  - *Perceive Status*
    On GUIs, the status is typically perceived through the visual sense, i.e. something needs to be displayed on the screen. The more prominently this signal is shown against other UI elements, the faster it can be found and seen. Rich interaction features such as animations can help make status elements more salient.
  - *Understand Status*
    Once data has been perceived, the next step is to understand it. On a user interface, for example, a status message may be clearly visible, but its meaning may be ambiguous to the user.
  - *Compare w/ Goal*
    Once the meaning of a current status is understood, it has to be put in relation to the envisioned goal. How near or far am I away from that goal?

  - *Plan Operation*
    In order to bridge the gap between current and envisioned state a strategy has to be developed.
  - *Anticipate Outcome*
    An operation is selected based on the expectation that it will reduce the gap between current and envisioned state. So one way to select from several operations is to anticipate its outcome, i.e. its effectiveness.
- Operate:
  - Initiate Operation
    Even if the current state has been correctly assessed and an effective operation has been planned, it can be a challenge to start executing on that operation. In the area of user interfaces an example is a user who knows that s/he has to increment a value on a virtual gauge, but doesn't know how to do it.
  - Monitor Operation
    For operations more complex than a single click, the actual execution has to be monitored. This can be viewed as a nested T.O.T.E. within the Operate T.O.T.E. unit. For instance, at the end of a drag & drop action the user has to monitor whether the dragged element is on the target drop zone, otherwise the action won't be successful.
- Test:
  - see above –
- Exit:
  - Re-Orient
    After successfully ending a T.O.T.E. sequence the attention has to be shifted to the next activity. On user interfaces this may involve scanning the screen for other information, other visual targets that allow initiating the next T.O.T.E. sequence.

The drill-down into the sub-processes of T.O.T.E. units helps to understand the complexity behind seemingly simple user activities within an application. It also helps understanding how rich interactivity can support the human-system interaction. This will be explained in the next sections.

2.2 Nuanced Feedback

For human-system interaction that includes direct manipulation of UI elements, the analysis of T.O.T.E. above showed that during longer interactions like for Drag & Drop the user needs to monitor the operation, i.e. making sure that the object to be relocated is dragged on to a valid target zone. De-composing a Drag & Drop action into more atomic events leads to what Bill Scott (Scott, 2009) calls "interesting moments", incl.:

- Mouse hover
- Mouse down
- Drag initiated
- Drag leaves original location
- Drag re-enters original location

- Drag enter valid target
- Drag exits valid target
- etc.

Scott lists 15 events for Drag & Drop. With the user input being the same (mouse down – dragging of element to target zone – mouse up), rich interactivity provides several alternatives on how the system output during the Drag & Drop action can be presented (see Figure 1):

- The appearance of the drop zone can be changed.
- The appearance of the drag object itself can be changed.
- The appearance of the mouse cursor can be changed.
- Any combination of the above.

The last bullet calls attention to the fact that combinations of rich behavior can be detrimental for the user experience as too many different stimuli have to be perceived and understood at the same time.

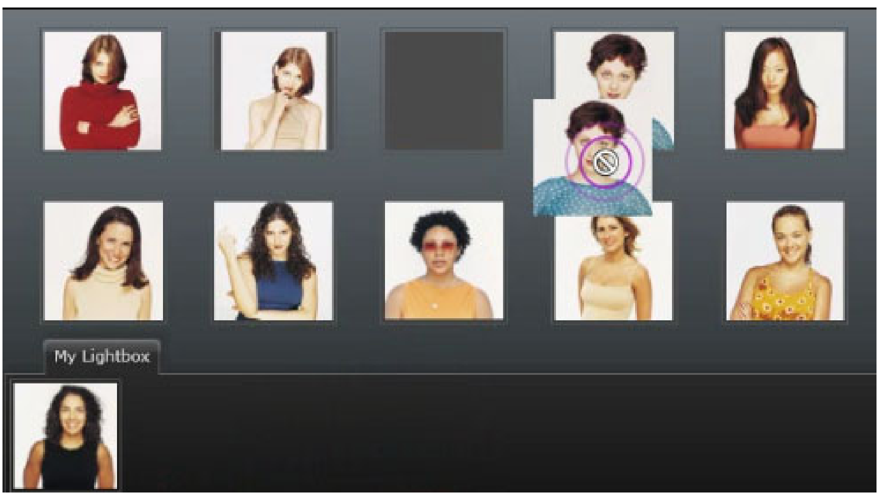


Fig. 1. Rich system behavior for Drag & Drop: mouse cursor indicating that the target zone has not been reached yet

Traditionally, in UI Specs standard mouse events have been documented, for example what happens on Mouse hover, Mouse down, Mouse up, etc. With the capabilities of designing Scott's interesting moments, the documentation effort is significantly higher as with rich interactivity different system outputs can be defined for different interesting moments. Even when these are documented in a specification document, consumers of that specification may find it hard to envision what the rich interaction looks and feels like. Building mock-ups for demonstration conveys rich concepts better than static words and images, but the effort to create them is significantly higher as well.

2.3 Keeping Users in the Same Screen Context

One of the biggest challenges in user interface design is to shorten navigation paths and de-cluttering screens – both at the same time. Rich interaction concepts allow keeping users on the same screen longer by providing means that go beyond the traditional tabs, accordions and blinds that only have to states (on/off, expanded/collapsed, shown/hidden, etc.):

- Microsoft uses a composition model in WPF and Silverlight that allows placing any UI element into any other UI element. For example, a rich tool tip on a grid cell displaying last year's total sales in the U.S. can host a visualization of the regional sales numbers that make up the total number – this contextual information is provided right where it is needed without requiring the user to navigate somewhere else in order to find it.
- User experience patterns such as in-line editing or drill-down charts enable the user to keep working on an item without leaving the current screen context. Navigation steps are minimized and consequently the perceived complexity of an application is reduced.
- Oftentimes, data objects and functional objects that are only infrequently needed are positioned in fly-out menus that can be brought into view on demand. Rich interaction patterns can help reduce the effort in terms of user for accessing these "additional" spaces through more seamless ways of access (interesting moments, see above).

2.4 Education

Interaction concepts can be challenging for users to learn and understand. Rich interactivity can be used for educating users. For example, when a user clicks on a UI element which in turn is replaced by another object (e.g. drill-down charts, see above) animations showing how the new object superimposes on the initial object help communicating that switch that would otherwise be binary and instant. In that respect animations can support the concept of location consistency, i.e. that each object has its place on the UI, even though it may not be visible at all times.

Another example of how a rich interaction feature contributes to a better understanding of a UI is the so-called UX pattern "Magnetism". It specifically helps making affordances explicit. Magnetism denominates the behavior of an interactive UI element when the mouse pointer comes close. Imagine a fly-out menu that is collapsed with only few pixels sticking out of the side of the screen in order to maximize screen real estate. Once the mouse pointer gets close, the menu will move inward by a couple of pixels. By that, it is communicated that the element is interactive (since it reacts to the mouse pointer being close by). Furthermore, its intended behavior is indicated through the movement.

An important aspect of user education is feedback. Especially on web sites the user oftentimes does not perceive and understand feedback, e.g. after input validation has returned an error on a form. This can be explained by change blindness (Simons, Ambinder, 2005). Through rich interactivity the feedback message or alert can be asynchronously loaded and be animated to draw attention.

2.5 Entertainment

Entertainment is an integral part of many web sites to help attract and retain site visitors. The convergence of desktop and web worlds has also brought with it the deployment of entertaining features into line-of-business applications. For these applications it feels counter-intuitive to incorporate entertainment, as productivity is one of the biggest goals. And still, entertainment serves its purpose and richness adds to its quality. Entertainment in line-of-business applications constitutes all those features of a front-end that serve the purpose of diverting the users and holding their attention. Entertainment features are not required for the accomplishment of user objectives. In most cases, the goal is to bridge waiting time for the user during the initial start-up of the application, during times when more data and content is loaded as part of using the application, and during processing times. In all these situations the user is put into a passive mode where he or she has to wait before relevant information is displayed or inputs can be made.

The start-up of an application, for example, can be lengthy. During this time, so-called splash screens are typically shown, initial screens containing information like product name, company logo, version information, etc. In recent years splash screens have become richer. Animations and multimedia see to it that the perceived waiting time is shortened. The richness also generates an expectation towards the interaction quality of the product once it has been loaded.

3 Outlook

User interface design is constantly evolving and advancing. New concepts and new technologies build the basis for novel interactive systems. The role of user experience design remains the same – to center the design of a software product on the needs, capabilities and limitations of the users. User-centered design processes and methods remain crucial in order to exploit the strengths and to mitigate the weaknesses of new technologies.

References

1. ISO 9241-210, Ergonomics of human-system interaction – Part 210: Human-centered design for interactive systems. Geneva: International Organization for Standardization (2010)
2. Miller, G.A., Galanter, E., Pribram, K.A.: Plans and the structure of behavior. Adams-Bannister-Cox, New York (1986)
3. Scott, B., Neil, T.: Designing Web Interfaces: Principles and Patterns for Rich Interactions. O‘Reilly Media, Cambridge (2009)
4. Simons, D.J., Ambinder, M.S.: Change blindness: Theory and consequences. Current Directions in Psychological Science 14(1), 44–48 (2005)

Axiomatic Design for Biometric Icons

Sheau-Farn Max Liang and Chien-Tsen Lin

Department of Industrial Engineering and Management, National Taipei University of Technology, 1, Sec. 3, Chung-Hsiao E. Rd., Taipei 10608, Taiwan, ROC
maxliang@ntut.edu.tw

Abstract. Well-designed icons should be visually distinctive and appropriately represent their intended meanings. Through the axiomatic design method, potential design issues about the recognition of a set of biometric icons for a fingerprint capture device have been identified. Two possible solutions, icon redesign and provision of procedure instruction, were proposed. A 2x2 ANOVA experiment was designed to test the main and interaction effects of the two solutions on eighty participants' icon recognition rate. Results showed that either redesigning the icons or providing the procedure instruction could significantly enhance the icon recognition rate. Furthermore, significant interaction effect indicated that the benefit of instruction was minor if redesigned icons were applied. From the perspective of human factors, icon redesign according to the axiomatic design principles may be a predominant solution.

Keywords: Axiomatic Design, Icon, Biometric Device, Fingerprint Capture.

1 Introduction

Icon is one of the most important components of a Graphical User Interface (GUI) due to its aesthetic attractiveness, possibility of rapid recognition, and potential of internationalization. An icon is usually not designed alone but together with other icons as a set. The perceived similarity between two icons is increased by the number of common features and decreased by the number of distinctive features [1]. Confusion may occur when similar or common features shared among icons that represent different meanings [2]. Therefore, a well-designed icon should be visually distinctive and appropriately represents its intended meanings, that is, the discriminability and meaningfulness of icons [3]. In this study, these two key factors are achieved through the axiomatic design method [4, 5].

Axiomatic Design (AD) method 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 [6]. 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 [7]:

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 98–106, 2011.
© Springer-Verlag Berlin Heidelberg 2011

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 [8], 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 ≥ 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:

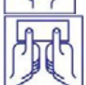
$$I_i = \log_2 \frac{1}{P_i} = -\log_2 P_i \tag{6}$$

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.

As part of the National Institute of Standard and Technology (NIST) biometric usability research, a set of eleven icons designed for biometric systems was analyzed according to the axiomatic design method. As listed in Table 1, these icons represent user action indications and device feedback when users interact with the device, such as (1) ready state, (2) fingerprinting on the platen, (3) wait/hold, (4) press more/less, (5) acceptable/unacceptable capture(s), and (7) exit or see guard.

Table 1. Eleven biometric icons and their twelve intended meanings

Icon	Intended Meanings	Icon	Intended Meanings	Icon	Intended Meanings
	(1) Ready state (2) Acceptable captures		Scan right-hand fingers		Wait/Hold
	Unacceptable captures		Scan left-hand fingers		Press more/less
	Acceptable scan		Scan thumbs		See guard
	Unacceptable scan		Exit		

2 Analysis

The purpose or functional requirements (FRs) of the biometric user interface is to instruct users to interact with the device and provide necessary feedback. Types of feedback includes (1) ready state, (2) acceptable captures, (3) unacceptable captures, (4) acceptable scan, and (5) unacceptable scan. On the other hand, types of user action indication includes (1) scan right-hand fingers, (2) scan left-hand fingers, (3) scan thumbs, (4) see guard, (5) exit, (6) press more/less, and (7) wait/hold. The hierarchy of functional requirements (FRs) is presented in Figure 1.

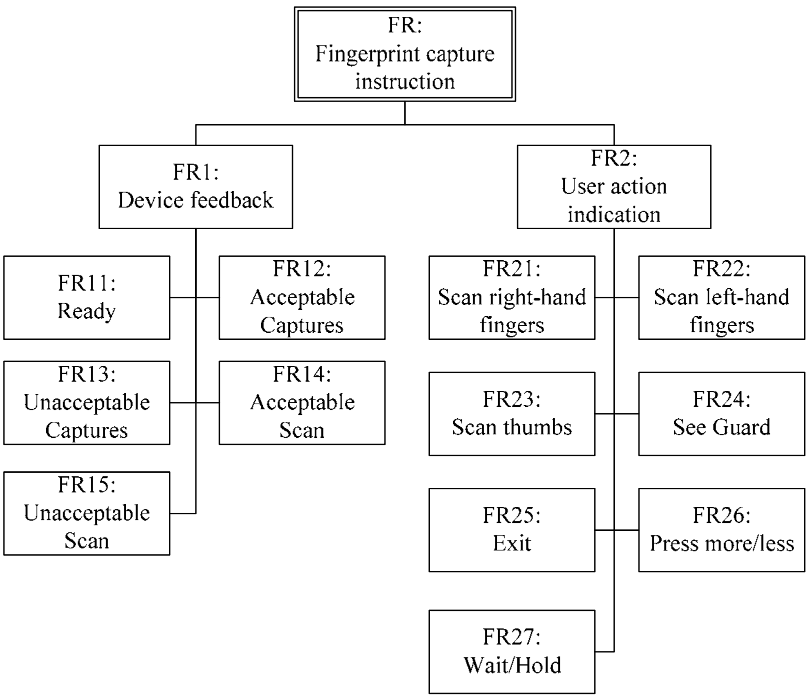


Fig. 1. Hierarchy of functional requirements (FRs)

Based on the original icon design, design parameters (DPs) of the biometric user interface were first categorized into (1) symbol, (2) pictogram, and (3) color contrast. DPs in the symbol category were (1) OK in a green circle, and (2) OK in a red prohibitive circle. DPs in the pictogram category were (1) two light bulbs, (2) a light bulb, (3) right hand on the platen, (4) left hand on the platen, (5) thumbs on the platen, (6) human figure and a guard figure, (7) human figure and a gate with an arrow sign, (8) finger on the surface, and (9) digits and a clock. DPs in the color contrast category were (1) blue-on-white, and (2) white-on-blue. The hierarchy of original design parameters (DPs) is presented in Figure 2.

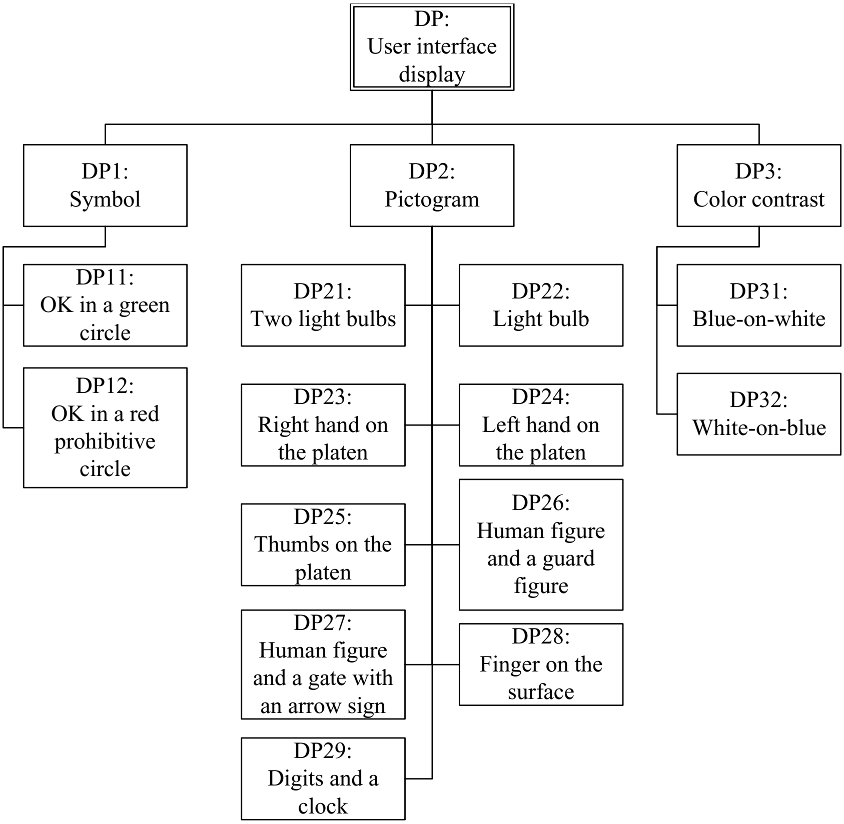


Fig. 2. Hierarchy of original design parameters (DPs)

A design matrix of twelve second-level FRs and thirteen second-level DPs was established as Eq. 7.

$$\begin{Bmatrix} FR11 \\ FR12 \\ FR13 \\ FR14 \\ FR15 \\ FR21 \\ FR22 \\ FR23 \\ FR24 \\ FR25 \\ FR26 \\ FR27 \end{Bmatrix} = \left[\begin{array}{cc|cc|ccccccc|cc} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}\right] \begin{Bmatrix} DP11 \\ DP12 \\ DP21 \\ DP22 \\ DP23 \\ DP24 \\ DP25 \\ DP26 \\ DP27 \\ DP28 \\ DP29 \\ DP31 \\ DP32 \end{Bmatrix} \quad (7)$$

The values of elements in the design matrix showed that the device feedback (FR1) and user action indication (FR2), two first-level FRs, could be well distinguished by their color contrast (DP3). The second-level FRs in the user action indication category could be further distinguished by their corresponded pictograms (DP2). That is, they were uncoupled design. However, the second-level FRs in the device feedback category and their associated DPs were coupled designs as indicated in the dotted square in the design matrix. This may confuse users in recognizing these icons. Therefore, five new icons presented in Table 2 were designed according to the axiomatic design principles.

Table 2. Original and new design of biometric icons

Original	New	Intended Meanings
		Ready state
		Acceptable captures
		Unacceptable captures
		Acceptable scan
		Unacceptable scan

As shown in Eq. 8, the new three second-level DPs in the pictogram category (DP21, 22, and 23) grouped the possible device feedback into three stages: (1) before, (2) during, and (3) after the fingerprint captures, whereas the new three second-level DPs in the symbol category (DP11, 12, and 13) grouped the possible device feedback into three states: (1) ready, (2) acceptable, and (3) unacceptable. It was assumed that the new design would improve the icon discriminability and meaningfulness, and ultimately, the icon recognition. An experiment was designed to test the hypothesis.

$$\begin{Bmatrix} FR11 \\ FR12 \\ FR13 \\ FR14 \\ FR15 \end{Bmatrix} = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \begin{Bmatrix} DP11: OK\ in\ a\ grey\ circle \\ DP12: OK\ in\ a\ green\ circle \\ DP13: OK\ in\ a\ red\ prohibitive\ circle \\ DP21: Light\ bulb\ (off) \\ DP22: Light\ bulb\ (on) \\ DP23: Index\ finger\ on\ the\ platen \end{Bmatrix} \tag{8}$$

3 Experiment

In the experiment, two independent variables were the icon design with two levels of new versus original design and the provision of the procedure instruction with two levels of with versus without instruction. The dependent variable was the icon recognition rate. A questionnaire was designed to list the icons and their intended meanings parallel with them but in a random order. Participants were asked to match the best meanings with each icon.

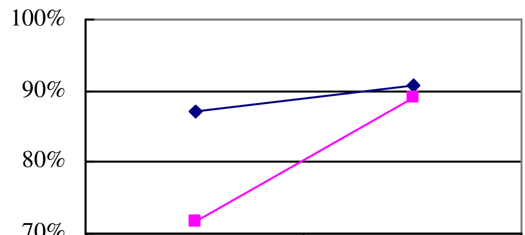
Eighty university students, fifty-six male and twenty-four female, were recruited as participants of the experiment. The mean value of their ages was twenty-four years old with the standard deviation of 1.3 years old. Participants were randomly divided into four groups for four experimental conditions to answer an icons-meanings matching questionnaire. Half participants answered the questionnaire with original icons, and the other half answered the questionnaire with new icons. Within each of these two groups, twenty participants were randomly selected to provide their fingerprint capture procedure instruction during answering the questionnaire. Figure 3 illustrate the procedure instruction with original icons. The icon recognition rate was then calculated for each participant as the number of his/her correct matches divided by the total number of matches.

Fig. 3. Fingerprint capture procedure instruction with original icons

4 Results

Results showed that the mean value (M) and standard deviation (SD) of the icon recognition rate for each condition was: (1) M = 71.7%, SD = 14.9% for original icons without instruction, (2) M = 89.2%, SD = 11.5% for original icons with instruction, (3) M = 87.1%, SD = 13.4% for new icons without instruction, and (4) M = 90.8%, SD = 11.8% for new icons with instruction, respectively.

A two-way ANOVA test was conducted for the main and interaction effects of the independent variables on the dependent variable. Results showed that the recognition rates of the two levels of the icon design (new vs. original) differed significantly, with $F(1, 76) = 13.45$, $p <.001$. A significant main effect of the provision of procedure instruction (with vs. without) was also found with $F(1, 76) = 8.69$, $p =.004$. That is, either redesigning the icons or providing the procedure instruction would significantly enhance the recognition rate. The interaction effect between the two independent variables was also significant, with $F(1, 76) = 5.63$, $p =.02$. Further post-hoc pair comparisons revealed that the improvement of recognition rate by applying both icon redesign and instruction was insignificant compared to the applications of either one solution. Figure 4 illustrates this interaction effect between the icon design and provision of procedure instruction.



	Without instruction	With instruction
New design	87.1%	90.8%
Original design	71.7%	89.2%

Fig. 4. Interaction effect of design and instruction

5 Conclusion

The axiomatic design method applied in this study demonstrates that it is a systematic and effective tool for analyzing and evaluating the design of icons. Design potential issues can be identified and solutions can be derived through the design matrix between the functional requirements (FRs) and design parameters (DPs). Furthermore, it is more cost-effective compared to those user-based methods that require the involvement of users. It is also more objective compared to those inspection-based methods that rely on the subjective judgment of evaluators.

The findings of this study show that user recognition on those biometric icons could be improved by the icon redesign according to the axiomatic design principles, or by the provision of procedure instruction. However, from the perspective of human factors, we think the resign is a superior solution than the instruction since the design issues have not been amended in the latter. Besides, user recognition on the icons could be improved by one-time icon redesign rather than providing extra procedure instruction every time. The application of the axiomatic design method on icon design should be able to provide relevant and useful suggestions for the design of public icons or symbols about not only "what" well-design icons should be but also "how" to design them.

Acknowledgments. This research is supported by the National Science Council of Taiwan (NSC 99-2221-E-027-073-MY2). Thanks to Dr. Yee-Yin Choong of the NIST for her coordination of the project.

References

1. Tversky, A.: Features of similarity. Psychological Review 84, 327–352 (1977)
2. Wickens, C.D., Lee, J.D., Liu, Y., Gordon Becker, S.E.: An Introduction of Human Factors Engineering, 2nd edn. Pearson Prentice Hall, Upper Saddle River (2004)
3. Liang, S.-F.M.: Analyzing icon design with axiomatic method: A case study of alarm icons in process control displays. Asian Journal of Ergonomics 7(1-2), 11–28 (2006)
4. 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)
5. 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)
6. Suh, N.P.: Axiomatic Design: Advances and Applications. Oxford University Press, New York (2001)
7. Suh, N.P.: The Principles of Design. Oxford University Press, New York (1990)
8. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1949)

A Rapid Prototyping Tool for Interactive Device Development

Mark Merlo[1,2] and Mark Bachman[1]

[1] University of California, Irvine, The Henry Samueli School of Engineering, University of California, Irvine, CA 92697-2700, USA
[2] Wildflower Technologies, Inc. 1300 S. Bristol St., Santa Ana, CA. 93704
mwmerlo@wildflowertechnologies.com, mbachman@uci.edu

Abstract. Designers need rapid prototyping tools that are embeddable, easily configured and can control a large range of accessories. Current prototyping tools fall short on these requirements by requiring one or more of the following: a tether to a computer, textual programming, and/or limited accessory control. To overcome the limitations of current tools, we have developed Buttercup, a standalone embedded sensor/effector controller that provides a high degree of customization for rapid prototyping interactive devices. The keys to the implementation of Buttercup are its hardware and firmware architecture. By building a system focused on sensor and effector control, the hardware can be small and inexpensive. The firmware utilizes a unique mapping system that lends itself to robust control over its accessories while allowing intuitive configuration by the user through a graphical user interface.

Keywords: Embedded controller, rapid prototyping, graphical programming, physical interactive system, effector controller, sensor controller.

1 Introduction

Designers are increasingly adding electronics and computers to products to increase their functionality and interactivity [1]. Rapid prototyping and the iterative design process are critical to the development of new products [2]. It is typically difficult or prohibitively expensive to incorporate form and function into a single prototype and, therefore, form and function are routinely divided during the development and prototyping of new devices. This difficulty is compounded by the time and cost that it takes to add or change features of a functional prototype [3]. This leads to fewer functional prototypes during the prototyping phase and the design process suffers from fewer iterations.

Tools that enable functional prototypes are of great value to the designer. From the designer's perspective, the ideal prototyping tool would be embeddable and small, be easy to configure with no textual programming, and offer control over a large range of sensors and effectors (i.e. devices that modify the environment or provide stimulus to humans; examples include lights, sounds, vibrators, motors, and heaters). Embeddability and large range of accessories are important to make the tool as universal as possible.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 107–113, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Ease of configuration without programming is critical since designers generally do not have a programming background [3].

Other groups have identified these needs and have developed systems to aid the designer in the prototyping process. Despite these efforts, with the exception of the Buttercup platform presented here, no prototyping tool satisfies the three needs listed above (i.e., embeddable, simple configuration and control of large range of sensors and effectors).

Many prototyping tools have focused on the information appliance [4-6] due to its popularity and ubiquity in today's culture. An information appliance is a computer designed to perform a specific function; examples include mobile phones, global positioning systems (GPS) devices and MP3 players [7]. Even though it is not explicitly stated in the definition of information appliances, these devices almost always interact with the user through a visual display, audio or both. Since both visual displays and audio require high bandwidth and fast computing power, almost all prototyping tools use a traditional PC to control these outputs and thereby tether, either wired or wirelessly, the prototyping tools to a computer. This tether usually prevents the device from being embedded or limits the range the device is functional. Some examples of these prototyping tools include d.tools [5], Thumbtacks [8], and DisplayObjects [6].

Phidgets [9] has the ability to control many types of sensors and effectors, but still requires a tether to a PC. The Arduino and iStuff Mobile [10] can work independent of a computer and lend themselves nicely to being embedded. The Arduino can also control many sensor and effector types, but requires textual programming to operate and configure (a severe drawback to less program-savvy designers). iStuff Mobile suffers from lack of effector support and though it can be visually programmed, the type of visual programming is aimed towards programmers.

To overcome the shortcomings of current rapid prototyping tools, we have developed the Buttercup platform, an embeddable sensor/effector controller that can be intuitively configured visually.

2 Platform Overview

The Buttercup is a standalone embedded sensor/effector controller that provides a high degree of customization for rapid prototyping interactive systems. Fig. 1 shows the Buttercup controller along with some sensors and effectors that are compatible with the system. In its most basic operation, sensor data is read by the Buttercup controller which uses a user-defined response curve to generate control signals for an effector.

The block diagram of a generic system setup in Fig. 2 shows the basic system components and typical data flow. The four basic subsystems of this platform are: the Buttercup controller, a sensor, an effector and power. There is also an optional connection to a computer via USB that is used to configure the device using a graphical user interface (GUI) and to monitor its operating parameters in real-time. After the designer has configured the device, the parameters can be saved to non-volatile memory and the Buttercup controller can be disconnected from the computer. When power is supplied to the Buttercup device (e.g. from batteries), the parameters that were saved are loaded and the system will operate independent of a computer.

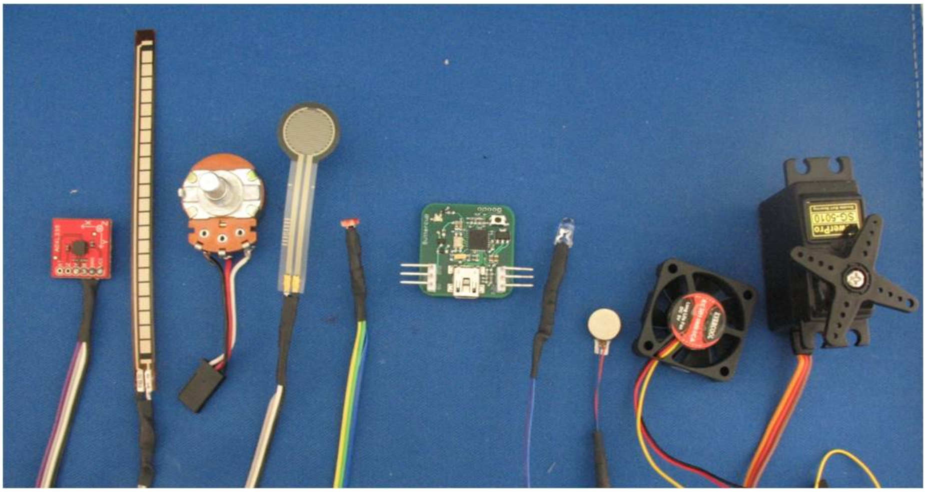


Fig. 1. Common components used in the Buttercup platform. From left to right: accelerometer, flex sensor, potentiometer, pressure sensor, light sensor, Buttercup controller (center), LED, vibrator, fan, servo-motor.

3 Platform Details

3.1 Buttercup Controller

The Buttercup controller is a 26 mm x 26 mm device that is used to read sensor data and control an effector. The main hardware components of the Buttercup controller are a PIC18F24J50 microcontroller, a 5 V to 3.3 V voltage regulator, two crystals (32.768 kHz and 12 MHz), a mini-USB connector and headers for a sensor and an effector. The interfaces to the sensor and actuator operate on 5 V power and the microcontroller and all the supporting components operate on 3.3 V power. The analog-to-digital converter (ADC) on the microcontroller is 8-bits and therefore can record 256 discrete sensor values. During duty cycle control, the pulse width modulated (PWM) output of the microcontroller operates at 40 kHz with a resolution of 8-bits, or 256 discrete duty cycles. This PWM frequency was selected to be above the upper threshold of human hearing (~20 kHz) and therefore limit any audible sound that might be generated when driving certain effectors. In servo mode, the PWM frequency is 50 Hz and the pulse width ranges from 1.2 ms to 1.8 ms (1.5 ms centered) at a resolution of 1 μs. This control signal is typical of most servo-motors.

3.2 Sensor

The three pin sensor connector from the Buttercup controller provides ground, 5 V power and analog signal in. The analog input signal should range between 0 and 5 V. The sensor response is not required to be linear since the designer can correct for nonlinear sensors in the configuration of the Buttercup controller, but the sensor should have an impedance of 10 kOhms or less to provide sufficient current to the ADC of the microcontroller.

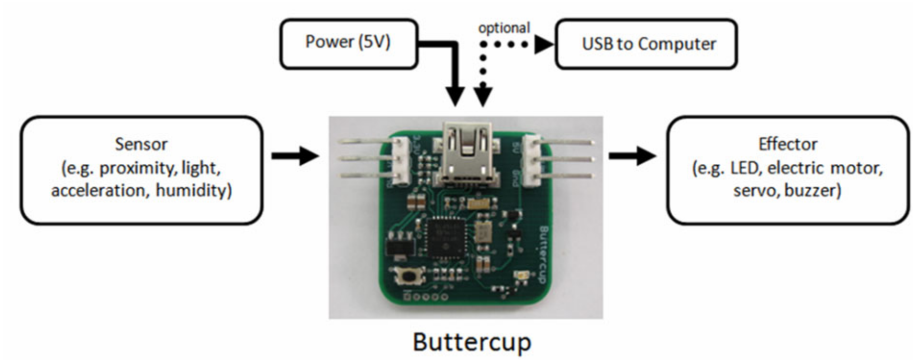


Fig. 2. Block diagram of a generic interactive system utilizing the Buttercup controller

3.3 Effector

The effector can be any device that is controllable by a PWM signal. This includes motors, LEDs, vibrators and speakers. The effector can also be controllable by pulse width duration, such as a servo-motor, instead of the typical duty cycle controlled devices mentioned above. The three pin effector connector from the Buttercup controller provides ground, 5 V power and PWM signal out.

3.4 Power

The operating supply power can range from 4.5 V to 5.5 V. Power is provided to the Buttercup controller through the USB connector from either the 5 V USB bus power or from batteries during standalone operation. Power to the effector and the sensor are provided by their respective connectors on the Buttercup module.

4 Platform Architecture

There are three main elements to Buttercup's architecture that enable its functionality and versatility as shown in Fig. 3. The entire system update period is 1 ms, which is

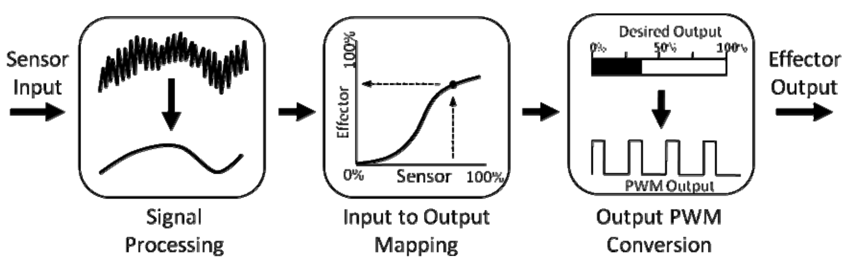


Fig. 3. Diagram of Buttercup architecture: Sensor data is processed from left to right to determine the proper effector control signal

fast enough for any human interface device since the discrete steps are not perceivable and the system latency is not noticeable.

The first element, the signal processing step, reduces the noise of the analog sensor signal by oversampling the sensor data at 1 kHz. The data sampled is averaged over the system update period of 1 ms, thereby averaging 10 readings for each update period. This signal processing reduces the noise from the sensor and eliminates most high frequency signals without affecting the response of the system.

The second element, sensor-to-effector mapping, enables the designer to arbitrarily control the effector's response to the sensor signal. By providing complete control over the sensor-to-effector responsive curve, the designer has the ability to fine-tune every aspect of the effector's response throughout the entire sensor range. This includes simple adjustments such as limiting the maximum or minimum effector output to more complex responses such as linearizing the sensor output or even reversing the effector response as the sensor signal increases (i.e., the effector goes from 0% to 100% back to 0% when the sensor signal goes from 0% to 100%). This sensor-to-effector response is the same curve that is labeled C in Fig. 4.

In the firmware this response curve is stored as a 256 element lookup table, with the array index being the raw sensor value as determined by the signal processing step. A lookup table was selected over interpolation for several reasons. First, since the sensor data is only 8-bits and the PWM output resolution is 8-bits, a lookup is not prohibitively large, even for a microcontroller. Second, computation time is much smaller than other methods, such as interpolation, since the microcontroller does not have a floating point coprocessor. Finally, the lookup table gives designers complete control over the response curve and does not limit them to finite control of just a few points.

The third element, output PWM conversion, adjusts the output PWM signal to the effector based on the desired effector output. Depending on the effector, different PWM signals are needed to control it throughout its full range. For example, to turn an LED from 0% (fully off) to (100%) fully on, the PWM signal must be adjusted from 0% duty cycle to 100% duty cycle. In contrast, to move a common servomotor over its full range, the PWM signal must be adjusted from a 1.2 ms pulse width (6% duty cycle) to a 1.8 ms pulse width (9% duty cycle). The output PWM conversion element of the Buttercup provides the flexibility to control a large number of different effectors.

5 User Interface

Although the Buttercup platform can operate independently of a PC, a PC is required to configure the device. The user interface for configuring the Buttercup platform is shown in Fig. 4. The graphical display, labeled C, gives the designer complete control over the sensor-to-effector response curve (black lines). In this figure, the preset cosine curve has been loaded. Each point on this curve can be dragged and dropped, allowing fine control over the system's response. Within this same display, the blue crosshairs provide the real-time operating point of the system. This feedback allows the user to quickly see how the system is responding and where adjustments need to be made. This real-time response is especially useful for determining the operating parameters of a sensor since the designer can see the real-time sensor output.

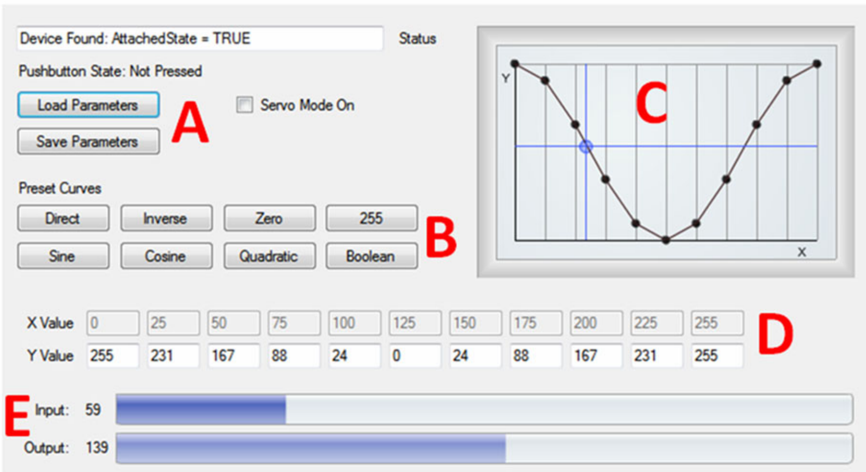


Fig. 4. Screenshot of the user interface: The large red letters are not part of the actual interface and are used only to highlight important aspects of the interface. A: Buttons used to load and save data to the Buttercup. B: Common curves can be loaded with these buttons. C: Graphical display of the sensor-to-effector response curve. Blue crosshairs provides real-time data of the current operating mode. D: Text boxes show numerical values of sensor-to-effector response curve. E: Real-time data of the sensor input signal and the effector output control signal.

The meter bars on the bottom, labeled E, provide real-time values of the sensor reading and the desired effector output. The Servo Mode On checkbox near label A allows the designer to select a PWM output that is compatible with a servo, or when unselected, typical PWM duty cycle control is used.

The Load Parameters button near label A loads the current response curve in volatile memory of the Buttercup controller. The Save Parameters button puts the response curve into non-volatile memory which is loaded on startup and allows independent operation.

This user interface and the low level USB communication was written in C# with the exception of the response curve and meters bars, which were written in Flash.

6 Conclusion

Reducing the time between design iterations is critical to speeding up the prototyping process. Currently, there is no other embeddable system that provides a convenient method for rapidly prototyping interactive systems with a large range of accessories. Buttercup fills this need by providing an effective platform for prototyping devices that utilize sensors and effectors. The Buttercup platform gives designers complete control over their system's response. The parameters of the Buttercup platform are controlled with an intuitive user interface that allows the designer to adjust critical parameters of the interactive system. This enables the designer to quickly and effectively test new design concepts. Future work includes expanding on the types of sensors and effectors that Buttercup supports, adding support for multiple sensors and effectors, and increasing the functionality to include different types of control schemes.

References

1. Gill, S.: Six Challenges Facing User-oriented Industrial Design. The Design Journal 12 (2009)
2. Ahmed, S.M.Z.: A user-centred design and evaluation of IR interfaces. Journal of Librarianship and Information Science 38, 157–172 (2006)
3. Avrahami, D., Hudson, S.: Forming interactivity: a tool for rapid prototyping of physical interactive products. In: 4th Conference on Designing Interactive, pp. 141–146 (2002)
4. Carlson, J., Han, R., Lao, S., Narayan, C., Sanghani, S.: Rapid prototyping of mobile input devices using wireless sensor nodes. In: Proceedings of the Fifth IEEE Workshop on Mobile Computing Systems & Applications, pp. 21–29 (2003)
5. Hartmann, B., Klemmer, S.R., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., Gee, J.: Reflective physical prototyping through integrated design, test, and analysis. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology - UIST 2006, p. 299 (2006)
6. Akaoka, E., Ginn, T., Vertegaal, R.: DisplayObjects: Prototyping functional physical interfaces on 3D styrofoam, paper or cardboard models. In: Proceedings of the Fourth International Conference on Tangible Embedded and Embodied Interaction, p. 49 (2010)
7. Gill, S., Loudon, G., Woolley, A., Hare, J., Walker, D., Dix, A., Ellis, D.R.: Rapid development of tangible interactive appliances: achieving the fidelity/time balance. International Journal of Arts and Technology 1, 309 (2008)
8. Hudson, S.E., Mankoff, J.: Rapid construction of functioning physical interfaces from cardboard, thumbtacks, tin foil and masking tape. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, p. 298. ACM, New York (2006)
9. Greenberg, S., Fitchett, C.: Phidgets: easy development of physical interfaces through physical widgets. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, pp. 209–218. ACM, New York (2001)
10. Ballagas, R., Memon, F., Reiners, R., Borchers, J.: iStuff mobile: rapidly prototyping new mobile phone interfaces for ubiquitous computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI 2007, p. 1107. Citeseer (2007)

Key Requirements for Integrating Usability Engineering and Software Engineering

Karsten Nebe[1] and Volker Paelke[2]

[1] University of Paderborn, C-LAB, Fuerstenallee 11, 33102 Paderborn, Germany
Karsten.Nebe@c-lab.de
[2] Institut de Geomàtica, Av. Carl Friedrich Gauss, 11, 08860 Castelldefels, Spain
Volker.Paelke@ideg.es

Abstract. To improve the integration between Software Engineering (SE) and Usability Engineering (UE) this paper identifies areas of overlap and develops proposals for their integration. The focus is on key requirements that were derived using semi-structured interviews and questionnaires. The principles and activities in the standards ISO 13407 and ISO/PAS 18152 were concretized to establish specific quality aspects. The identified requirements provide a foundation for systematic modification of existing development processes with established best practices from both disciplines.

Keywords: Integration, Assessment, Standards DIN EN ISO 13407 and ISO/PAS 18152, Process Models, Process Definition.

1 Introduction

Software development processes aim to produce products with a guaranteed quality level, where quality can refer to a variety of aspects including reliability and maintainability from a technical perspective as well as usability and adequateness from a usage perspective. Software Engineering (SE) has developed SE models as systematic approaches to address technical quality goals during development. SE models define detailed activities, the sequence in which these activities have to be performed, and the resulting deliverables and are meant to unify working processes [13]. To address the quality goals from the usage perspective the domain of Usability Engineering (UE) has also developed a variety of methods and processes. The development of high quality software requires to combine the goals and methods of SE and UE in a way that allows systematic implementation while considering all quality factors as well as costs and time. A key question is how this integration can be realized in practice. The introduction of novel development processes that address all quality requirements and integrate methods from both disciplines is problematic. In most development contexts a modification of existing SE processes is desired. The challenge is to provide systematic support for the assessment of existing SE processes and guidelines for their extension with usability oriented activities.

In the study described here the first aim was to identify an appropriate level of abstraction at which such guidance can be provided that is neither restricted to a specific process but concrete enough to add value to the integration process. The second goal

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 114–120, 2011.
© Springer-Verlag Berlin Heidelberg 2011

was to provide concrete guidance through an assessment process at this level of abstraction.

Therefore the authors first analyzed various types of different integration approaches in order to identify the most promising level of abstraction for further integration strategies. The level of software engineering and usability standards is selected for closer inspections, i.e. in terms of assessment approaches. Previous work is summarized and the proceedings of an expert evaluation (interviews) are presented. The resulting requirements can be used first, as a basis for the definition of software development processes aiming at usable solutions; and second, as an enhancement to existing assessment standards. Then, first steps towards validation and application of these requirements are discussed.

2 Integration Approaches

In theory and praxis, a considerable number of integration approaches exist [15] (detailed analysis can be found at [10]). Some of these approaches tend to define common activities and artifacts for both SE and UE and to integrate these specific activities into the process of development. They aim at a 'soft integration‘ of UE aspects on a mutual basis, e.g. at interlinking relative results. Most approaches focus on a minimal organizational and structural transformation and/or change. For example, Schaffer [14] presents a method for the integration of UE activities, which is based on the evaluation of a concrete existing process.

Other integration approaches address the level of process definitions and process models. These aim to define pre-settings for the development and contain both a more concrete approach (focusing on the integration of UE activities in an already existing SE Models), and more fundamental aspects of process models (independently of any concrete SE Model). An example for approaches that integrate UE activities with existing SE Models is Pelka [12] who build on the V-Model as a base for the integration of UE activities.

A third group of integration approaches focuses at an even higher level of abstraction and describe organizational measures, principles, paradigms or meta-models. Those approaches are independent from any specific process model or activities but rather describe organizational measures, principles, paradigms or meta-models. As an example, Pawar [11] analyzed several activities of SE and UE in order to identify principal similarities as the basis for a framework for integration.

These approaches can be structured into three levels of abstraction [6]: The abstract level of standards, the level of process models and the operational process level. These are related in a hierarchy: standards define the overarching framework, process models describe systematic and traceable approaches within such a framework, and at the operational level the models are tailored to fit the specifics of an organization [6].

3 Process Assessments

This hierarchy exists in both disciplines software engineering and usability engineering and can be exploited for integration. At the level of standards general integration strategies can be defined that are applicable to a large number of existing development processes. For the purpose of integration we have focused on the SE standard ISO/IEC 12207 [2] and the UE standard ISO 13407 [1] . ISO/IEC 12207 defines a process framework for the development and management of software systems and identifies required processes and activities for the development-lifecycle. ISO 13407 defines four activities of human-centered design that should take place during system development.

An analysis [6, 7] shows that while some guidance can be derived from matching these standards, in practice more detailed and adequate criteria for the assessment are necessary to make objective statements about the ability of process models to create usable software and to derive actions for process improvement. We have therefore followed a process assessment approach that aims at the evaluation and optimization of existing processes in order to ensure technical quality and usability. Separate assessment approaches exist in SE and UE. ISO/IEC 15504 [3] describes a methodology and structure for assessing software development processes and defines requirements for the assessment, an assessment model and a reference model. For the domain of UE a similar approach exists with ISO/PAS 18152 [4]. To ensure usability within the development process there is a need for specific quality criteria for the assessment that integrates usability criteria.

The authors believe that there is such thing as a 'common understanding' in terms of what experts think of when they talk about UE and this is certainly represented by the definition of the human-centred-design process in the DIN EN ISO 13407. Although the definitions of base practices defined in the ISO/PAS 18152 are not considered as invalid they leave leeway for interpretation [9].

However, while this 'common understanding' seems true on a very abstract level, strong differences in how to implement these in practice can be expected. The key question therefore is not only what should be done, but rather how it can be assured that everything needed is being performed (or guaranteed) in order to gain a certain quality of a result, an activity or the process itself. In addition, the completeness and correctness of the base practices and human-centred design activities as defined in the ISO/PAS 18152 itself needed to be verified.

To establish such criteria based on current best practice the authors performed semi-structured interviews and questionnaires with six experts in the field of UE [10]. The experts were volunteers from an expert working-group on the integration of software engineering and usability engineering, the ISO committee TC 159 Ergonomics/SC4 as well as industry experts. A substantial part of the interviews referred explicitly to quality characteristics/aspects of the human-centered design activities of ISO 13407: 'context of use', 'user requirements', 'produce design solutions' and 'evaluation of use'. The goal was to identify what constitutes the quality of a certain activity from the experts' point of view and what kind of (potentially measurable) success and quality criteria exist that are relevant on a process level and subsequently for the implementation in practice.

As a result, overarching process- and quality characteristics were derived that led to statements about the relevance, the application and need of usability activities, methods and artifacts to be implemented in SE.

In summary, it can be said that the quality of the four activities essentially depends on the production and subsequent treatment of the result generated by each activity.

From the quality perspective it is less important how something is accomplished, but rather to guarantee the quality of the results. In order to answer the question what constitutes this quality, the analyzed statements of the experts regarding each activity have been analyzed and documented. In [9] the core characteristic (essence) of each activity is described, followed by requirements regarding the generation and treatment of content, a summary of (measurable) quality criteria and success characteristics, as well as a list of operational measures that can be used for the implementation in practice.

After this, the authors categorized and structured each single statement (473 altogether) by: an unique id; its origin (question concerning one activity: 'context of use', 'user requirements', 'produce design solutions' or 'evaluation of use'); and the references to ISO 13407 and/or ISO/PAS 18152 (questions aimed on the goals and base practices of the two standards). After that, the statements have been categorized whether they addresses quality aspects regarding the process, an activity, or deliverable. All those parameters have been used to perform several iterations of analysis. By doing this each statement of each single interview partner has been compared with the statements of the others in order to count the number of similarities in terms of the content. Those statements that reflect a similar issue have been combined to a general requirement. As a result, the more statements refer to one requirement the more relevant it has been rated. By doing this, the statements have been fused and formalized into 107 'requirements for development processes or process models' that define the demands of UE (see Table 1 for an excerpt, complete list will be online for publication shortly). At last, the requirements have been categorized according to their focus on: quality aspects, the process, activities or deliverables.

As a result the authors distinguish two distinct types of requirements: Compliancy and key requirements. Compliancy requirements represent the goals and base practices defined in the standards and refine them with the input of the experts from the interviews. Key requirements define core characteristics of the overall frameworks usability activities focusing on the quality of activities and results.

The beneficial use of the requirements is two-sided: First, as a basis for the definition of software development processes aiming at usable solutions; and second, as an enhancement to the assessment standard ISO/PAS 18152 and its assessment criteria.

4 Steps towards Validation and Application

Validation of process changes in complex processes like software development is difficult because the cost and complexity of real world development projects makes controlled experiments impossible. To validate the compliancy and key requirements we used two approaches.

In the first step we distributed a questionnaire to 13 usability experts (again involving experts from industry, academia and ISO standardization committees) in which they were asked to rate the requirements from their practical experience with regards to correctness (is this requirement correct) and relevance (how important is the requirement with regards to the overall goal of ensuring usability). Table 1 shows these rating for exemplary requirements. As a result, compliancy and key requirements were established that represent an evaluated knowledge basis for the development of usable products [8, 10].

Table 1. Examples of the requirements for the UE-activites 'context of use' (CoU), 'user requirements' (UR), 'produce design solutions' (PDS) and 'evaluation of use' (EoU) and the experts' rating in terms of correctness and relevance (in practice)

Nr	Activity	Requirement	Correctness	Relevance
2	CoU	Context-analysis is an integral part of the process.	Correct	Very high
17	CoU	The outcomes of the context analysis serve as the input for the next process step and the activity itself is anchored within the process model accordingly.	Correct	High
27	CoU	The characteristics of the intended users and their tasks, including user interaction with other users and other systems, are documented.	Correct	Very High
24	CoU	The analysis is focused on the original context of the users (their goals, tasks, characteristics of the tasks and the environment, etc.). The analysis is independent of any existing solution/implementation.	Correct	High
33	CoU	The context-information is based on facts and not an interpretation of any situation.	Sufficient	Medium
46	UR	A sufficient amount of user requirements are the basis for the next process step (PDS).	Correct	Very High
71	PDS	The development of solutions is carried out in collaboration with the development team.	Correct	Very High

In a second step we examined in how far the requirements are suitable to identify problems with existing processes as documented in literature. While a notable scarcity of such documented experiences exists there are some useful exceptions. In particular Jokela [5] has conducted and documented eleven process assessments within the last years and identified a number of common usability problems in industrial practice, e.g. Missing or Only Partial Usability Engineering Activities, Results of Usability Activities Not Considered, Quality Problems, Knowledge and Skills Problems, etc. By matching these problems identified by Jokela in 11 assessment projects with the compliancy and key requirements, we established that the majority of the reported problems are addressed by the key requirements. This indicates that common problems of UE in development practice could be avoided if the compliance and key requirements were considered during the definition of development processes.

The requirements could also be used as detailed criteria for assessing processes (or process models) leading to more specific results about weaknesses. In addition, more specific measures for improvement could be derived than in common assessment approaches (i.e. the ISO/PAS 18152).

The compliancy and key requirements therewith not only reflect basic demands to be applied on development processes they also address common problems of integrating SE and UE. By fulfilling these requirements a substantial contribution to the development of usable solutions could be made.

Altogether the requirements lead to the integration of SE and UE on all three levels of abstraction. They define goals and characteristics of activities, quality and success criteria as well as results that lead to the systematic development of usable products. This is reflected through the overarching framework and therewith on the level of standards, by the requirements' ability to be applied on the level of process models and the operational level in which the models are tailored to fit the specifics of an organization.

5 Conclusion and Outlook

This paper identifies compliancy and key requirements for the integration of usability engineering activities into software engineering processes. The 107 requirements define objectives and characteristics of activities as well as quality and success criteria for execution and were ranked and rated by usability engineering experts to provide an indication of their validity and relevance. The requirements are abstract enough to serve as assessment criteria for a wide range of processes but their application in an assessment process provides more specific guidance than generic guidelines. The requirements can be used to refine to the ISO/PAS 18152, a public available specification for a process assessment of human-system issues.

Currently the authors are analyzing the changes in the new version of the standard ISO 9241-210 (which replaced the ISO 13407) and are going to adapt the requirements where necessary. In addition the ISO/IEC 12207 is being operationalized in more detail (activities, tasks, artifacts and their relations) and are mapped to the compliancy and key requirements. The results will end in a new reference model for integrating software engineering and usability engineering likewise. The model can be used for process definition and assessments as well. A first expert based evaluation of the model is going to be finished very soon.

In future work the authors aim to apply the optimized assessment approach to additional process models, such as agile approaches, to establish their suitability beyond traditional SE processes.

References

1. DIN EN ISO 13407: Human-centered design processes for interactive systems, CEN - European Committee for Standardization, Brussels (1999)
2. ISO/IEC 12207: Information technology - Software life cycle processes, 2nd ed. 2008-02-01. ISO/IEC, Genf (2008)
3. ISO/IEC 15504: Information technology - Process Assessment - Part 1-5. ISO/IEC, Genf (2003-2006)
4. ISO/PAS 18152: Ergonomics of human-system interaction - Specification for the process assessment of human-system issues. ISO, Genf (2003)
5. Jokela, T.: Performance Rather than Capability Problems. Insights from Assessments of Usability Engineering Processes. In: Bomarius, F., Komi-Sirviö, S. (eds.) PROFES 2005. LNCS, vol. 3547, pp. 115–127. Springer, Heidelberg (2005)

6. Nebe, K., Zimmermann, D.: Suitability of Software Engineering Models for the Production of Usable Software. In: Gulliksen, J., Harning, M.B., van der Veer, G.C., Wesson, J. (eds.) EIS 2007. LNCS, vol. 4940, pp. 123–139. Springer, Heidelberg (2008a)
7. Nebe, K., Zimmermann, D.: Aspects of Integrating User Centered Design to Software Engineering Processes. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4550, pp. 194–203. Springer, Heidelberg (2007)
8. Nebe, K., Paelke, V.: Usability-Engineering-Requirements as a Basis for the Integration with Software Engineering. In: Jacko, J.A. (ed.) HCII 2009, Part I. LNCS, vol. 5610, pp. 652–659. Springer, Berlin (2009)
9. Nebe, K., Zimmermann, D., Paelke, V.: Integrating Software Engineering and Usability Engineering. In: Pinder, S. (ed.) Advances in Human-Computer Interaction, ch. 20, pp. 331–350. I-Tech Education and Publishing, Wien (2008)
10. Nebe, K.:: Integration von Usability Engineering und Software Engineering: Konformitäts- und Rahmenanforderungen zur Bewertung und Definition von Softwareentwicklungsprozessen, Doctoral Thesis, Shaker Verlag, Aachen (2009)
11. Pawar, S.A.: A Common Software Development Framework For Coordinating Usability Engineering and Software Engineering Activities. Master Thesis, Blacksburg, Virginia (2004)
12. Pelka, A.I.: Integration von SW- und Usability-Engineering am Beispiel des V-Modells. Diplomarbeit, Rheinische Friedrich-Wilhelms-Universität, Bonn (2001)
13. Rechenberger, P., Pomberger, G.: Informatik-Handbuch. Carl Hanser Verlag, München (2002)
14. Schaffer, E.: Institutionalization of usability: a step-by-step guide. Addison-Wesley; Pearson Education, Inc., Boston (2004)
15. Seffah, A., Desmarias, M.C., Metzker, E.: Human Centered Software Engineering, HCI, Usability and Software Engineering Integration: Present and Future. In: Seffah, A., Gulliksen, J., Desmarais, M.C. (eds.) Human-Centred Software Engineering – Integrating Usability in the Development Lifecycle, vol. 8. Springer, Berlin (2005)

Message Oriented Middleware for Flexible Wizard of Oz Experiments in HCI

Mirko Otto, Rafael Friesen, and Dietmar Rösner

Department of Knowledge Processing and Language Engineering
Otto-von-Guericke-University Magdeburg, Germany
{miotto,friesen,roesner}@ovgu.de

Abstract. Wizard of Oz (WOZ) systems and WOZ experiments are an important tool for basic and applied research in HCI. We report about using SEMAINE as a flexible component based middleware with a loose coupling of components as software infrastructure for WOZ experiments in human companion interaction. We focus on our experimental WOZ designs, their realisation within the SEMAINE framework and lessons learned from deploying the implemented solutions as the basis for ongoing controlled experiments with 120 subjects.

Keywords: Wizard of Oz, Companion Systems, Emotion, Multimodal, Message Oriented Middleware, SEMAINE.

1 Introduction

A Companion System (CS) can be described as *"an agent or 'presence' that stays with the user for long periods of time, developing a relationship and 'knowing' its owners preferences and wishes. The Companion communicates with the user primarily through speech, but also using other technologies such as touch screens and sensors."* [3]. Companion systems will assist their users in managing their daily life. An elderly person that may still live on his own with such a technical support is more likely envisaged as a companion user than a young professional. So support in mundane activities rather than in specialised business applications is of high relevance.

In order to explore companion systems, Wizard of Oz (WOZ) experiments are a usual approach [5,6,11]. The WOZ concept has a common usage in the fields of experimental psychology, human factors, ergonomics, linguistics, and usability engineering to describe a testing design methodology where an experimenter ("Wizard") simulates the behaviour of a theoretical intelligent computer application.

Different software components communicate with each other, thus the system needs an infrastructure that supports flexibility and reusability. Message Oriented Middleware (MOM) is a software infrastructure with the focus on sending and receiving messages between different systems [2]. In general MOM supplies a base for application interoperability in heterogenous and complex environments. Application Programming Interfaces (APIs) which are distributed across diverse

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 121–130, 2011.
© Springer-Verlag Berlin Heidelberg 2011

platforms and networks are typically provided by a MOM. This facilitates easy integration of components running on different operating systems and written in different programming languages. The SEMAINE API provides an abstraction layer over a MOM and provides the needed functionality for modular systems dealing with emotions [9].

In this paper we suggest a framework, with which WOZ experiments can be provided in flexible way.

This paper is organized as follows. We first introduce the Wizard of Oz paradigm in section 2. Section 3 describes our experimental design. In section 4 we present the architecture of the framework followed by an description of the implemented interface components in Section 5. Section 6 concludes the paper and provides some insights on our future work.

2 Wizard of Oz

Wizard of Oz (WOZ) systems are an important tool for basic and applied research in Human-Computer Interaction (HCI) and thus likewise for companion systems. This is especially true when WOZ experiments are interleaved with the development and implementation of additional system functionality, e. g. development and/or integration of autonomous and automatic components that replace functionality formerly provided by the wizard.

A WOZ system allows the observation of a user operating an apparently fully functioning system whose missing services are supplemented by a hidden wizard. The user is not aware of the presence of the wizard and is led to believe that the computer system is fully operational. The wizard observes the user through a dedicated computer system connected to the observed system over a network. When the user invokes a task that is not available in this observed system, the wizard simulates the effect of the task. Through the observation of users behavior, designers can identify users needs when accomplishing a particular set of relevant functions and evaluate the particular interface used to accomplish the functions. During a WOZ experiment, the exchanged data between the observed user and the wizard is recorded for further analysis [8].

3 Experimental WOZ Design

By definition a companion is with his user for a long period of time. So certain aspects of the human companion interaction will only be adequately dealt with in long term investigations with developed companion systems that users can carry with them and make use of in their daily life. For our experimental design we had to take into account that such developed systems do not yet exist. Even more some aspects of the interaction with a companion have to be investigated before such systems can be designed and implemented. In our experiment we investigate intentions ascribed to the companion system by users. Further description of this paradigm can be found in [7].

3.1 The Cover Story

Our cover story that is told to the subjects in the wizard of oz experiments takes these constraints into account: The subjects are told that they will run through the personalisation phase of a companion system – a phase that is a prerequisite for the system to be later adapted to the individual user characteristics and preferences – and that they therefore both have to answer a number of questions as well as to succesfully complete some tasks. Another aspect of this cover story is the resulting personal involvement of the subjects that is additionally supported by addressing the subject with his name and displaying the name as running head in various screens.

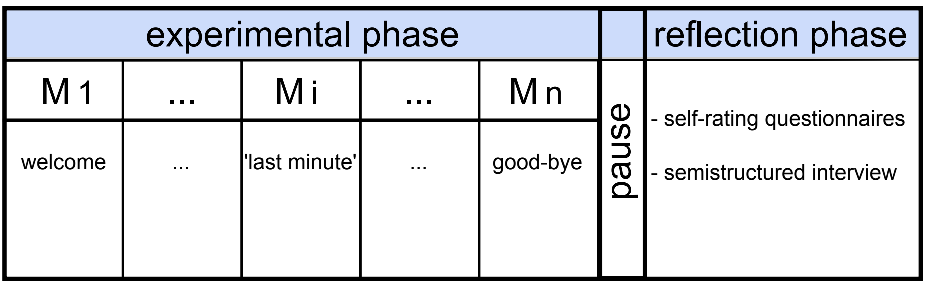


Fig. 1. Design of the WOZ experiment

The overall design of our WOZ experiment comprises an experimental and a reflection phase. The experimental phase was divided into different modules (c. f. Fig. 1). Each module stands for a part in the experimental phase and can employ different software components.

The module named 'last minute' was designed to investigate how users interact with a companion system in a mundane situation with the need for planning, replanning and strategy change. The subject interacts only vocally with the system.

3.2 Last Minute

The user is asked to imagine a situation in summer with most of his friends already in holidays when s/he as a surprise gets informed to have won a two weeks vacation. The prize includes as well the opportunity to select all necessary items for the travel from an online catalogue organized in a number of categories (e. g. coats, trousers and skirts, hats, underwear, sports equipment, etc.). For this selection fifteen minutes are available, because then the taxi to the airport will arrive and the packed suitcase will be available at check in.

4 Architecture of the Framework

The aim of our research was the development of an instrument for the realization of WOZ experiments, which fulfills the following requirements:

1. simple, easy and efficient development as well as integration of modules, software components
2. and distributed WOZ experiments across different locations.

The software for the WOZ experiments is component based and thus modular structured. The components communicate using a MOM which is specifically designed to integrate different applications or processes through messages being routed from publishers to subscribers. The aim of this "glue technology" is to glue together applications both within and across organizations, without having to reengineer individual components. An important advantage of a generic Message Oriented Middleware lies in its flexibility. The system architecture can be rearranged very easily – adding or removing a component consuming or creating a specific message type does not require any changes elsewhere in the architecture [2].

In the first implementation OAA (Open Agent Architecture) [4] was used as MOM but this platform was replaced by the SEMAINE API [9]. The SEMAINE API provides an abstraction layer over a MOM that allows the components to deal with messages in a type-specific way. The MOM currently used in the SEMAINE API is ActiveMQ[1] from the Apache project.

Our approach uses the SEMAINE API, which is an open source framework for building emotion-oriented systems and provides Java and C++ wrappers around a MOM. This framework allows the communication between components running on different operating systems via XML and other standardized formats. Different components can run on physically distinct hardware as long as they are connected within a network.

5 Implementation

The interface components of the module 'last minute' are shown in Fig. 2. This 'last minute' module uses a control interface for the wizard ("wizard interface"), an visual user interface for the subject and a text-to-speech system for the speech synthesis ("subject interface"). The experiment is controlled by the Wizard via the control interface. From here the Wizard can select text blocks for the speech synthesis and control the visual operational sequence on the monitor of the test person [5].

An example interaction:

1. The subject says he would like to pack items (e. g. "I want to pack two t-shirts").
2. The Wizard understands the utterance, selects the equivalent term on the monitor and thus confirms the selection.

[1] http://activemq.apache.org/

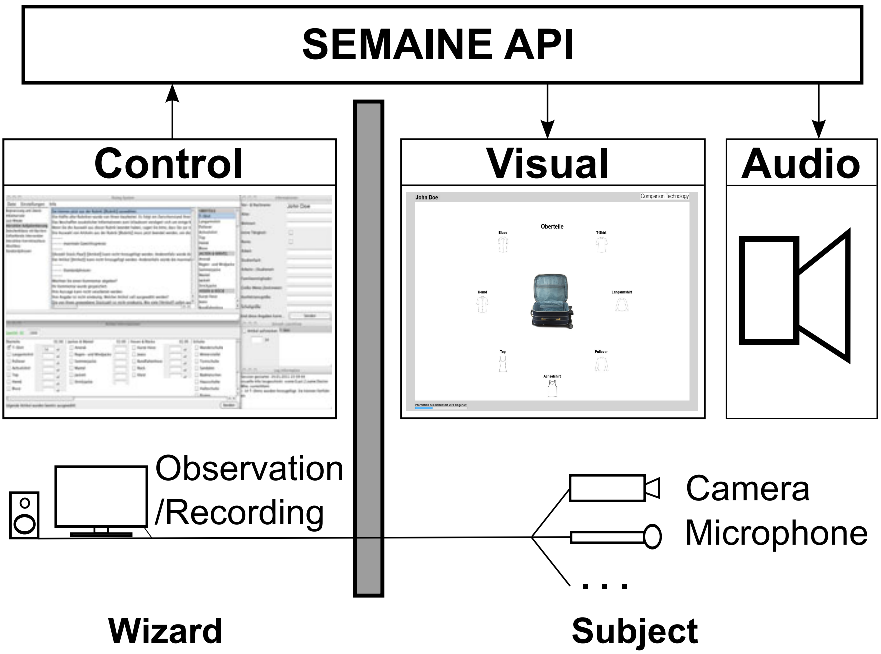


Fig. 2. Interfaces of the 'last minute' module

3. The SEMAINE API sends this information to the visual user interface and to the speech synthesis component.
4. The visual user interface displays the new information and the speech synthesis component generates the respective language output (e. g. "two t-shirts were added").

The WOZ system used is based on three major components integrated using the SEMAINE API. These are the control interface component of the Wizard and the visual user interface and speech output component of the test person. In the following we will describe these implemented software components in more detail.

5.1 Wizard Control Interface

The wizard control component and its interface is shown in Fig. 3. It provides typical features like a configuration menu and a log window. The frames for the different tasks are arranged as follows.

The dialog control frame contains the selection of modules of the WOZ experiment (1), predefined text patterns (2), the item selection (3) and at the bottom a text field for free text input. For each module of the WOZ experiment selected all predefined text patterns are loaded and displayed. Some text patterns need additional information (e. g. which item should be packed or which catalogue section to load). After selecting a text pattern and – optional – an item the sentence

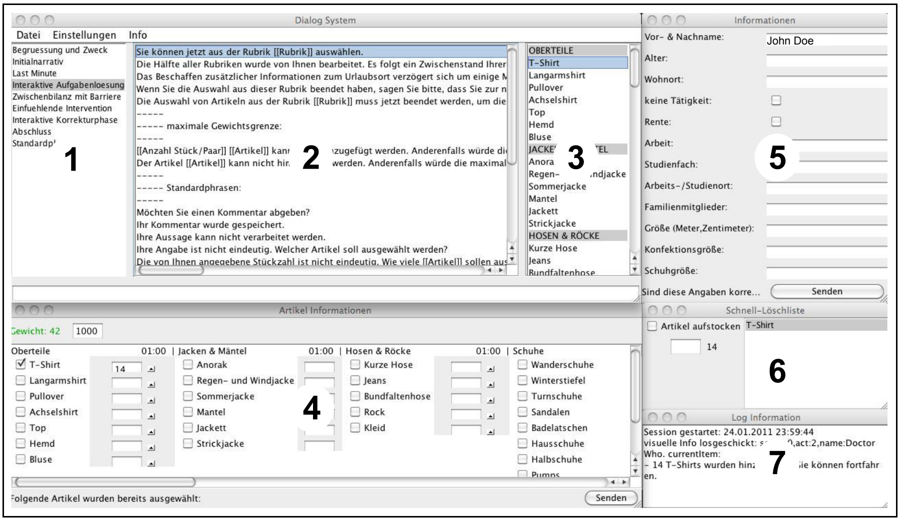


Fig. 3. Wizard Control Component

can be send to the other components by pressing 'enter' or with a double-click. If a sentence is needed but can not be found fast enough or is not available in the text pattern list, it can be typed into the free text field (providing auto completion) and sent. The Article information frame (4) provides an overview of all items already in the suitcase and their actual weight. It can also be used to pack items. The Information frame (5) provides the subjects personal information and is used in the personalization stage of the experiment. The Fast-Erase frame (6) shows all packed items in alphabetical order and is used to unpack items or change their amount quickly. The Log frame (7) shows a log of all events and sent messages to support the observation of the subjects screen. The Frames 1 - 6 support fast handling and provide easy sending functionality and hotkeys.

5.2 Visual User Interface

The visual user interface is started seperately but is controlled by messages sent from the wizard control component. It visualizes information about the experiment and the selection menus during interaction (c. f. Fig 4) as well as the subjects name on top and additional information about the session state. With this component the subject is able to observe his actions on the screen. The visual user interface is written in JavaFX[2].

5.3 Speech Component

The Speech Component receives messages send from the wizard control component and sends the messages to a MARY TTS[3] server to synthesise these

[2] http://javafx.com/
[3] http://mary.dfki.de/

Fig. 4. Visual User Interface: a selection menu

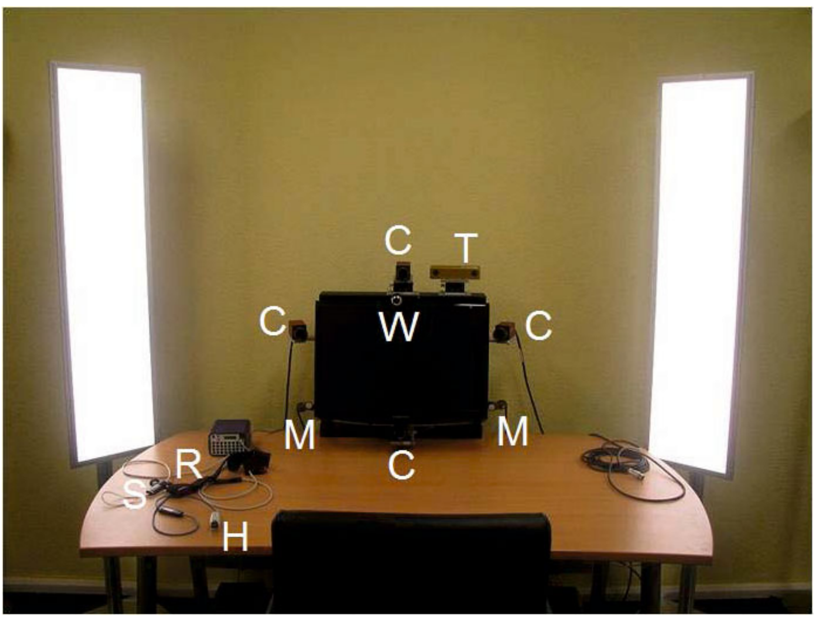


Fig. 5. The experimental hardware setting. C=High resolution camera, H=Heart beat clip, M=Microphone, R=Respiration belt, S=Skin conductance clip, T=Stereo camera, W=Observation webcam. Not in the picture: Headwear microphone.

messages. The MARY TTS platform is an open source and modular architecture for building text-to-speech systems [10]. It supports various pronounciation features of synthesised speech and will be employed to simulate emotionally sounding voice (e. g. friendly, sad, angry, . . .) in the ongoing improvement of the system.

5.4 Experimental Setting

The environment for the subjects (c. f. Fig. 5) is as follows: The subjects sit comfortable on a chair at a desk. They are told to lean back and are instructed not to move the left hand because the skin conductance and heart beat are measured at the left fingers. For respiration recording the subjects get an elastic belt around the chest. Sound is recorded with two microphones next to the screen and one headwear microphone. Video is recorded with one observation webcam, four high resolution cameras and one stereo camera. Illumination comes from two light panels next to the desk and three light panels at the ceiling. The window of the room is shut, so no daylight (constant illumination) and no noise comes in.

6 Conclusions and Future Work

In this paper we have proposed a software design for Wizard of Oz experiments that uses the SEMAINE API as the platform for the components' communication. This design allows flexibility for adding new components and exchange and reuse of components. The implemented software is used to perform experiments with 120 subjects and will be used – with different component constellations – for further experiments.

Automating parts of the WOZ experiment is work in progress. The architecture for automating the spoken dialogue between system and user is designed as follows. The software is divided into four SEMAINE components.

- One component connects a commercial speech recognizer (ASR) and text to speech (TTS) software[4] to the SEMAINE system. This component will from now on be referred to as the connector component.
- The container component acts as a storage for items which the user has packed into his luggage.
- A third component is called the control component and its task is to send commands concerning the insertion or removal of luggage items into the container, as well as keeping track of the active category for selection and notifying other components of which items are currently available for packing/removal.
- The last component is the grammar component based on Grammatical Framwork (GF) [1], which has two tasks. The first task is parsing utterances based on GF recognized by the ASR and transforming them into commands related

[4] http://www.nuance.com/

to packing/removing items from the luggage. These commands are then sent to the control component for further processing and, if applicable, to the container component, too. The second task of the grammar component is the reception and processing of status updates of the container. These status updates reflect successful operations inside the container. After such a status update the grammar component transforms the confirmation message into a natural language sentence, and sends this linearization to the text to speech component.

In addition to automating functionality of the wizard, future work includes the development of a formal description for the software components and experimental modules, e. g. by XML. Using this description several experimental settings (e. g. with more and different experimental modules, interfaces and components) can be created easily and with high flexibility.

Acknowledgements

The work reported is performed within the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG)[5]. The responsibility for the content of this paper lies with the authors.

References

1. Angelov, K., Bringert, B., Ranta, A.: PGF: A Portable Run-time Format for Type-theoretical Grammars. J. of Logic, Lang. and Inf. 19, 201–228 (2010), http://dx.doi.org/10.1007/s10849-009-9112-y
2. Banavar, G., Chandra, T., Strom, R., Sturman, D.: A Case for Message Oriented Middleware. In: Jayanti, P. (ed.) DISC 1999. LNCS, vol. 1693, pp. 1–17. Springer, Heidelberg (1999)
3. Catizone, R., Worgan, S.F., Wilks, Y., Dingli, A., Cheng, W.: A world-hybrid approach to a conversational Companion for reminiscing about images. In: Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues, pp. 157–168. John Benjamins, Amsterdam (2010)
4. Cheyer, A., Martin, D.: The Open Agent Architecture. Autonomous Agents and Multi-Agent Systems 4(1-2), 143–148 (2001), http://dx.doi.org/10.1023/A:1010091302035
5. Kelley, J.F.: An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. ACM Trans. Inf. Syst. 2(1), 26–41 (1984)
6. Legát, M., Grůber, M., Ircing, P.: Wizard of Oz Data Collection for the Czech Senior Companion Dialogue System. In: Fourth International Workshop on Human-Computer Conversation, pp. 1–4. University of Sheffield, Sheffield (2008)
7. Rösner, D., Friesen, R., Otto, M., Lange, J., Haase, M., Frommer, J.: Intentionality in Interacting with Companion Systems – an Empirical Approach. In: Jacko, J.A. (ed.) HCII 2011. LNCS, vol. 6761. Springer, Heidelberg (2011)

[5] http://www.sfb-trr-62.de/

8. Salber, D., Coutaz, J.: Applying the Wizard of Oz Technique to the Study of Multimodal Systems. In: Bass, L.J., Unger, C., Gornostaev, J. (eds.) EWHCI 1993. LNCS, vol. 753, pp. 219–230. Springer, Heidelberg (1993)
9. Schröder, M.: The SEMAINE API: Towards a Standards-Based Framework for Building Emotion-Oriented Systems. In: Advances in Human-Computer Interaction, vol. 21, pp. 319–406 (2010)
10. Schröder, M., Pammi, S.C., Türk, O.: Multilingual MARY TTS participation in the Blizzard Challenge 2009. In: Proceedings of Blizzard Challenge 2009. o.A. (2009)
11. Webb, N., Benyon, D., Bradley, J., Hansen, P., Mival, O.: Wizard of Oz Experiments for a Companion Dialogue System: Eliciting Companionable Conversation. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010). European Language Resources Association, ELRA (2010)

Design and Rich Application Frameworks

Kevin H. Richardson

Electronic Ink
1 South Broad St., 19th Floor, Philadelphia, PA 19107, USA
krichardson@electronicink.com

Abstract. With the advent of rich application frameworks like Flash and Silverlight as well as the increased exposure to interaction models they make possible (does anyone want a mobile device that doesn't have an iPhone-like interface?) it isn't difficult to imagine that usability and design professionals may be feeling a little vulnerable. After all, until recently, usability and design professionals were the last, best hope in the face of early web design, business systems left over from the 80s and clunky mobile phone menus. We helped create an environment in which users expected more (at least on the web). Today, however, developers have at their disposal an arsenal of tools designed to provide users with experiences that take advantage of asynchronous server calls, high-definition multimedia and slick, natural-feeling interactions. Have we been relegated to the role of usability testing? Surely, our profession has more to offer. The good news is that our role is the same as it has ever been. Like any platform or technology, rich application frameworks are the medium through which design is expressed. As such, they are no different from any previous platform that was ready to revolutionize the manner in which people interact with information, the world or each other. They are the tools through which researchers, designers and technologists enable users to complete tasks and make decisions. They are the paint and canvas, the clay and plastic molds, with which we bring our designs to life. A well-designed system is the result of a well-defined design process. That process includes the expertise of an interdisciplinary team with individual backgrounds in graphic design, fine art, architecture, cognitive psychology, anthropology, human-computer interaction, and other fields. This kind of design team has the training and experience to bridge the gap between business, technology and human requirements. They (we!) practice a design process that is mindful of the features, functions and legacy systems that must be somehow united, implemented and maintained. They are equally mindful of who will be using these systems (from motorcycle enthusiasts to financial analysts, from students to CEOs), their experiences and mental models, where the systems will be used (from hospital emergency rooms to living rooms and executive boardrooms) and what they need from technology to improve rather than impede outcomes. If, at any point, the user must wrestle with the interface, then research and design have failed. Our job, therefore, remains one of understanding the ways in which users need to have information presented to them, the ways in which they need to interact with it and the decisions they must make. The capabilities made available via rich internet applications provide a larger toolset from which to choose in order to meet these requirements.

Keywords: design, design process, research, usability.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 131–135, 2011.
© Springer-Verlag Berlin Heidelberg 2011

1 The Design Process

From the very conception of an interactive digital tool, everyone involved in the design and development process must understand the audience for whom they are designing. They must establish methods to gain that understanding. Researchers, designers, and developers must collaborate openly and freely to evaluate all decisions in terms of their understanding of the user. This is design, and it is precisely this kind of product design methodology that can be overlooked in the desire to create the type of interaction models made possible by rich application frameworks. Industry is too often led by technological innovation rather than by an understanding of human beings, their behaviors, their needs or their requirements.

The design process is ultimately research-based and highly iterative. It consists of gathering requirements from end users and stakeholders (both business and technology). The focus must be on creating exploratory models of how the user and system interact and on quickly testing the form of the design. Through incremental cycles of evaluation and refinement, the system is ensured to have been shaped accurately—not solely on the opinions of design professionals, but also on the needs, expectations and desires of the end-user audience. With the end user as a participant in the design effort, stakeholders are assured that the design is accurate and lasting.

Many companies think that by including richly interactive components to the user experience that they are somehow supporting users and following a creative design process. This may not be the case. I hope to demonstrate here why the design process is a prerequisite to the successful implementation of any designed experience.

1.1 Inform

Every effort begins with a thorough understanding of the user. The designers, with backgrounds and training in the fields of traditional design, psychology, anthropology, and human-computer interaction (among others) answer the following questions:

- *Who will use this software?* Is it an Electronic Medical Records (EMR) system with multiple physician specialties? A trade security application used by highly trained analysts? A public facing website used by the general public?
- *What tasks are users trying to perform?* Expectations of how tasks should be completed will not change to suit ill-designed software. A successful system allows users to perform tasks in natural and intuitive ways. Failed systems require users to adapt and change their behavior to meet the system demands. You can be certain that a sudden, increased need for "training" means that the system has failed the users.
- *In what context will the system be used?* A poorly designed system requires users to dedicate a (large) portion of their attention to understanding and navigating the interface rather than to their task. In critical situations, the amount of attention (and tolerance) available for deciphering the software decreases exponentially.
- *What expectations and experiences will users bring with them?* Users bring with them many expectations, experiences and preconceived notions of how things should work. It is important to keep in mind that the goals users have for any new

system are likely to be quite different from the goals of those responsible for purchasing, installing and maintaining it.

- What actions do users need to take as a result of this interaction? What decisions need to be made? Note that these questions, above all others, are what enable a design to be both useful and innovative. The form of the design is still a long way off at this point and it's still too early to tell whether users need the type of richly interactive models afforded by Flash or Silverlight.
- The quality of this analysis directly affects the outcome of the design process. Collectively, the information gained in this phase is vital to the design of any system destined to address the varied goals, strategies and requirements of users.

1.2 Discover

The design team next works with users and stakeholders to conceptualize the system, examining what will best meet both user needs and business goals. Research professionals on the design team seek to validate and discover the characteristics that best suit the new system, its audiences and their experience. Within a framework of existing systems, artifacts, business and technical requirements, the team considers limitations and defines the expectations against which a successful design will be measured.

1.3 Design

The Design phase is characterized as the iterative portion of the design process. Through collaboration during this phase, all participants (design team members, users, and stakeholders) are able to provide input and feedback that can influence design decisions. Using this information, designers can create novel and enduring designs that push the boundaries of existing systems while still enabling users to interact with information in natural and intuitive ways. And because these design prototypes can be produced and tested with users quickly with little to no impact on development, the design team has the freedom to explore innovative solutions. It is only now that specific interaction models (and their inherent technical requirements) can be considered. In outline form, the steps are simple:

1. What do users need to accomplish?
2. What form of data visualization/presentation best allows users to focus on information rather than translating data into information?
3. What tools, models and technology support the required interaction and data visualization model?

2 Isn’t Usability Enough?

Traditional usability firms (or usability groups within large companies) tend to focus on evaluation, and their design process typically ends at the Discover phase with workflows and application screens created by researchers. For companies that tout themselves as “Human Factors” or “Usability,” the **goal** is to have research and data dictate design. After all, isn’t a research-based interface what we’re after? In this environment, the role of design is often an afterthought and always under-used. While traditional usability firms (and the software development industry in general) have recognized the need for designers, it has typically been either to create nice looking marketing materials or to make the designs created by researchers "look pretty."

The reader can be forgiven if he or she has imagined me to be a (disgruntled) designer looking for a little respect. Not so. As a researcher with a classical background in experimental psychology (Ph.D. in Cognitive Psychology, 1991, The State University of New York at Buffalo) I have seen the field of ergonomics/human factors engineering/usability/design research develop and mature. However, in the 20 years I've being doing usability research, I've seen (and been party to) a great many applications that, while usable, were not innovative, inspiring, beautiful, or lasting regardless of the interaction models made possible by the technology du jour. Why? Because we were following a research process (or, dare I say, a technology-driven process) rather than a **design** process.

3 Design for Industry

Traditionally, the field of software development has looked to the contributions of two types of individuals to ensure that applications meet the needs of users. These individuals are the Business Analyst and the Subject Matter Expert. Many industries continue to rely on these job roles to gather, develop and coordinate application requirements. Indeed, the claims made by industry that their applications are guaranteed to be usable because they have business analysts (collecting business/functional requirements) and experts (representing users) on their staff are both commonplace and misleading.

Unfortunately, the fact is that no matter how well intentioned, these individuals are neither trained nor qualified to design a system that is usable, innovative and supportive. In order to understand what is missing from this process, you need only look at what these two job roles are expected to bring to the development process.

3.1 Business Analysts

The role of the Business Analyst (BA) is to gather functional requirements for the application from the point of the view of the business. In order to provide structure and guidance for the development team, these requirements are incorporated into "use cases." A use case is a detailed description of what the application needs to present to the user and what data the user needs to provide in order to accomplish a task. These use cases represent a view from the system's point of view—what does the system need to do for a task to be accomplished. For example, one might define the process of collecting patient information in terms of the fields on a screen that must be completed. However, this view of design ignores completely the needs of the user and the demands of the situation. It isn't difficult to see how this task might require differing designs to account for the different contexts of a clinician's office, an a trading floor, or a boardroom.

This is not to minimize the role of Business Analysts—it is extremely important to understand the needs of the business. However, accurate documentation of business requirements counts for little if users thumb their noses at the resulting application. Millions of dollars, years of development and crowds of dissatisfied users at companies all over the world are an unfortunate testament to this method of traditional software "design."

3.2 Subject Matter Experts

As a way of addressing the *problem* of users and their stubborn refusal to use systems that do not support them, industry has inserted the role of Subject Matter Expert (SME) into the mix.

In the medical industry, for example, this means making sure that software development firms employ clinicians (MD, RN, etc.) whose role is to describe what the application needs to do, what data is required and how it should be presented. While this might certainly be a step in the right direction, these individuals are fundamentally like any other user group—they are not skilled design professionals and their appointment to this role in the design process is completely flawed.

It is flawed because the job of the SME is specifically to act **in place of** users. Experts are, at best, providing what they believe to be user requirements or, at worst, their own requirements. The point is that they are not providing user requirements. In addition, Subject Matter Experts are also asked to make decisions regarding design. We should be clear on this point; experts are not qualified to design any more than designers are qualified to be subject matter experts. My apologies to the experts put in this role. With this in mind, users (and purchasers of new applications) should be wary of the marketing phrase, "Designed by Expert X for Experts Y." Perhaps, as an industry, we should start insisting on "Designed by Designers for Experts!"

4 What All This Means

Good design doesn't just happen. It is not a checklist of requirements, though it must certainly support the tasks that users need to accomplish. It is not a passing grade on a usability test; all the functionality in the world is meaningless if users can't or won't use the application. It is not a slick, natural-feeling interaction model. Most of all, good design is not "making it look pretty," though a well-designed application should certainly be expected to look the part. And neither business analysts nor Subject Matter Experts can guarantee good design.

So what is *good* design? Design is a research-based, highly iterative **process** with a focus on exploring different models of the user-system interaction. It consists of gathering requirements from end users and business and technology stakeholders coupled with fast, early and iterative usability testing to determine the ultimate form of the design. Perhaps this design incorporates highly interactive models of the sort made possible by today's rich application frameworks. Perhaps not. Whatever the form, it must support a user's requirement to get something done. This ensures that the system is shaped not only by the opinions of design professionals, but also on the needs, expectations and requirements of users.

Enhancing Personas with Their Main Scenarios

Alícia Valls, Muriel Garreta-Domingo, and Marta López

Universitat Oberta de Catalunya
Av. Tibidabo 47
08035 Barcelona, Spain
{avallssa,murielgd,mlopezreye}@uoc.edu

Abstract. The methodology of "personas" is a well-known procedure and very often used for characterizing target users in user-centered design projects. Very often, personas are represented by a picture. This paper introduces a different approach to present the results of a user analysis study with the goal to increase assimilation and comprehension of a project's personas. We believe that creating and developing a new concept enhancing "personas" with their context, giving less importance to the image of the person with a silhouette and adding an image of their main scenarios can make more impact in our organizations. To verify our idea we conducted a study to evaluate which format - "personas" with a picture and "personas" with their contexts – worked better.

Keywords: Personas, Persona usage, Context, Scenarios, User research, Design methods.

1 Introduction

There are many user-centered design methodologies with an important focus on communication and on how to convey information about target users to everyone with an impact on a product design. Among these methods is the technique of creating personas; introduced by Cooper [1].

Although practitioners find personas beneficial for communicating with stakeholders, guiding design decisions and evaluating design ideas [2], it is also true that "personas have unfortunately become more of a check-off item than a useful tool, and many personas get put on the shelf once they are written", as stated by Olsen [3].

Our own experience shows that user-centered design experts and even the Marketing department are very familiar with the personas (Jordi and Martina) that represent our main user archetypes. Nevertheless, this information has not yet downed on other team members. By creating a playful output of these personas and switching the weight into their context, we wanted to evaluate if the information is better conveyed and lasts longer.

2 Common Representations of Personas

Cooper [1] defines Personas as hypothetical archetypes, or "stand-ins" for actual users that drive the decision making for interface design projects. Personas are not real

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 136–141, 2011.
© Springer-Verlag Berlin Heidelberg 2011

people, but they represent real people throughout the design process. They are not "made up"; they are discovered as a by-product of the investigative process. Although personas are imaginary, they are defined with significant rigor and precision. Names and personal details are made up for personas to make them more realistic. Personas are defined by their goals. Interfaces are built to satisfy the personas' needs and goals.

Using the personas in user-centered design processes has huge benefits:

- They help team members share a specific, consistent understanding of various audience groups. Data about the groups can be placed in a proper context and can be understood and remembered in coherent stories.
- Proposed solutions can be guided by how well they meet the needs of individual user personas. Features can be prioritized based on how well they address the needs of one or more personas.
- Provide a human "face" so as to focus empathy on the persons represented by demographics.

However, in order to get these benefits, personas need to be presented and introduced to the design teams in effective ways. As a result, ideally, communicating personas should be a multifaceted, multimodal, and ongoing task that progressively discloses more and more information about them [4].

Posters, flyers, handouts, squeeze toys, beer glasses, mouse pads and even actors [5] have been used to introduce personas to the rest of the team and stakeholders. John Pruitt and Tamara Adlin [6] talk about three basic categories that describe the goals of the materials support: buzz generators, comparison facilitators and enriches. Despite all these different outputs, there is, however, consistency over the idea that a persona is originally composed of text and a picture representing the user [2].

3 A Different Approach to Present Personas

Our rationale behind designing and creating a new artifact and a new way of showing the personas information is:

1. Although a picture is worth more than a thousand words, we wanted to avoid the usage of a face. Images of people introduce biased notions that are often unstated and mostly unconscious. They rely too much on stereotypes and might shadow all the other information related to that persona.
2. Combining the concepts of personas and scenarios: we want to provide these user archetypes with a "real life" by means of an object and an image that would introduce them and give more importance to the context of their daily lives. The idea is not to provide and detailed scenario but deepen the information on each persona. The situations are the same for the two personas and present them at home, at work and commuting.
3. Foster understanding of our users by showing their main contexts. With the increase of mobile devices, we have seen a diversity of new environments from which our users interact with our product.
4. Create a playful and more permanent artifact to hold all this information.

4 Our Distance Learners' Personas

Over the years, especially since 2006, several studies have been conducted at the Universitat Oberta de Catalunya (Open University of Catalonia, UOC) – a completely online university - in order to know its students. These studies have taken into account the average UOC student types (mostly workers aged between 30 and 40 years) as well as users with accessibility issues, students in their commuting moments or extreme users (younger than 25 and older than 65). Several user-centered methodologies both qualitative and quantitative have been used in these studies: in-depth interviews, user tests, surveys, focus groups, contextual inquiries.

After analyzing all this information, we concluded that the UOC had two main personas: one persona that studies for pleasure, as a hobby; and another that pursues a degree to improve his career. We decided to call the two people Martina (the Hobby) and Jordi (the Executive).

With the two personas defined, we produced two different outputs in order to compare which worked better in terms of persona comprehension and persona information retention.

One output was a more common presentation of a persona, with its picture and a description on a sheet of paper. The other artifact – a cube that introduced the scenario-based persona - showed the same text but instead of the picture of a persona, there were photos of 3 daily situations for each. The personas were represented by silhouettes.

5 Personas vs. Scenario-Based Personas

The two artifacts – paper and cube - were presented to 20 people that work at the Office of Learning Technologies; the department in charge of designing, developing and maintaining the virtual campus at the UOC.

Ten people had the sheet of paper with the explanation of the individual scenarios together with the image of the persona (a more typical way to present the persona). The other 10 had the two cubes with the scenario-based personas.

After having some time to read the artifacts, each participant had to answer an online questionnaire with questions about the two personas, Jordi and Martina.

Fig. 1. Paper and cube artifacts

5.1 First Test

Persona: Martina (the Hobby)

From the 11 questions of the questionnaire: 5 questions had the same result with the cube and paper and 6 questions had better results with the cube.

Persona: Jordi (the Executive)

From the 11 questions that comprised the questionnaire: 4 questions had the same result with the paper and the cube and 6 questions had better results with the cube; 1 question was better answered with the paper.

First test		
Persona: Martina	answer comes from	
questions:	paper	cube
How old is Martina?	90%	100%
Is she married?	90%	100%
Does she have children?	100%	100%
Where does she work?	90%	100%
Why is she studying?	100%	100%
She has harried for finish her studies?	100%	100%
When she usually study?	70%	80%
She studies during her way to work?	80%	100%
Does she enter the UOC in her office?	70%	100%
Is she interested in the new technologies applied for learning?	60%	60%
What is Martina's motto?	100%	100%

First test		
Persona: Jordi	answer comes from	
questions:	paper	cube
How old is Jordi?	100%	100%
Is he married?	70%	90%
Does he have children?	100%	100%
Where does he work?	90%	100%
Why is he studying?	100%	100%
He has harried for finish his studies?	100%	90%
When he usually study?	80%	90%
He studies during his way to work?	50%	60%
Does he enter the UOC in his office?	60%	80%
Is he interested in the new technologies applied for learning?	50%	60%
What is Jordi's motto?	100%	100%

(green)	Best percentage
(orange)	Equal percentage

Fig. 2. First test results

Second test		
Persona: Martina	answer comes from	
questions:	paper	cube
How old is Martina?	60%	100%
Is she married?	80%	100%
Does she have children?	100%	100%
Where does she work?	80%	80%
Why is she studying?	100%	100%
She has harried for finish her studies?	100%	100%
When she usually study?	50%	60%
She studies during her way to work?	70%	90%
Does she enter the UOC in her office?	60%	70%
Is she is interested in the new technologies applied for learning?	40%	40%
What is Martina's motto?	100%	100%

Second test		
Persona: Jordi	answer comes from	
questions:	paper	cube
How old is Jordi?	60%	100%
Is he married?	70%	90%
Does he have children?	100%	100%
Where does he work?	80%	90%
Why is he studying?	100%	100%
He has harried for finish his studies?	100%	90%
When he usually study?	70%	90%
He studies during his way to work?	30%	70%
Does he enter the UOC in his office?	80%	80%
Is he interested in the new technologies applied for learning?	40%	60%
What is Jordi's motto?	100%	100%

(green)	Best percentage
(orange)	Equal percentage

Fig. 3. Second test results

5.2 Second Test

After a month, and in order to check out which of the two artifacts was better retained and therefore which one was more useful to ensure that team members remember the two personas, the same users were asked to answer the same questionnaire.

The results of the second test were:

Persona: Martina (the Hobby)
From the 11 questions: 4 questions had the same result with the paper and cube and 6 questions had better results with the cube; 1 question was best with the paper.

Persona: Jordi (the Executive)
From the 11 questions that comprised the questionnaire: 3 questions had the same result with the paper and the cube and 7 questions had better results with the cube; 1 question was best with the paper.

6 Conclusion

Overall and given the fact that between the two tests a month had passed, we can say that in general both artifacts had a good retention result. However, in the case of the test with the paper, the results were lower in the second test. The results obtained by the cube are more stable, as we can see in figure 4.

A part from the retention results, the participants that did the test with the cube showed greater enjoyment. The feedback from the users also confirmed that the images (paper format) transmitted more information than the one explicitly stated and which can introduce biases in the idea people have of the personas: "Jordi is posh and rich because it is what the picture shows".

martina			
First test		Second test	
answer comes from		answer comes from	
paper	cube	paper	cube
90%	100%	↓ 60%	= 100%
90%	100%	↓ 80%	= 100%
100%	100%	= 100%	= 100%
90%	100%	↓ 80%	↓ 80%
100%	100%	= 100%	= 100%
100%	100%	= 100%	= 100%
70%	80%	↓ 50%	↓ 60%
80%	100%	↓ 70%	↓ 90%
70%	100%	↓ 60%	↓ 70%
60%	60%	↓ 40%	↓ 40%
100%	100%	= 100%	= 100%

jordi			
First test		Second test	
answer comes from		answer comes from	
paper	cube	paper	cube
100%	100%	↓ 60%	= 100%
70%	90%	= 70%	= 90%
100%	100%	= 100%	= 100%
90%	100%	↓ 80%	↓ 90%
100%	100%	= 100%	= 100%
100%	90%	= 100%	= 90%
80%	90%	↓ 70%	= 90%
50%	60%	↓ 30%	↑ 70%
60%	80%	↑ 80%	= 80%
50%	60%	↓ 40%	= 60%
100%	100%	= 100%	= 100%

paper		cube	
=	40%	=	68%
↓	55%	↓	27%
↑	5%	↑	5%

Fig. 4. Results obtained from the cube

Taking into consideration the test results and the feedback from the users, we can conclude that with the scenario-based persona we: 1) reduced the bias that a face can generate using the silhouettes instead; 2) combined the concepts of personas and scenarios; and 4) created a playful and more permanent artifact to hold all this information.

However, we believe that we should further explore if the scenario based-persona fosters understanding of our users thanks to showing their main contexts. We have tested that it improves the retention of the information but we would like to evaluate as well if it helps improve the design work in order to make it more user-oriented.

In sum, our goal is to keep working on finding ways to ensure that everyone with an impact on the product design knows the personas and uses them as a guide to make design decisions. Besides this new persona format, we also developed a user-centered design game [7] and periodically have sessions to approximate UCD to the team. At the same time that we keep on running user studies to get to know our students from different perspectives.

References

1. Cooper, A.: The inmates are running the asylum. Macmillan, Basingstoke (1999)
2. Chang, Y., Lim, Y., Stolterman, E.: Personas: From Theory to Practices. In: Proc. Abstracts Nordi CHI 2008, pp. 439–442. ACM Press, New York (2008)
3. Olsen, G. Making Personas More Powerful: Details to Drive Strategic and Tactical Design. BoxesandArrows (2004), `http://www.boxesandarrows.com/view/making_personas_more_powerful_details_to_drive_strategic_and_tactical_design`
4. Pruitt, J., Grudin, J.: Personas: practice and theory. In: Proc. Abstracts DUX 2003, pp. 1–15. ACM Press, New York (2003)
5. Traci, L.: Personas: Explorations in Developing a Deep and Dimensioned Character. UXmatters (August 23, 2010) (published), `http://www.uxmatters.com/mt/archives/2010/08/personas-explorations-in-developing-a-deep-and-dimensioned-character.php`
6. Pruitt, P., Adlin, T.: The Persona Lifecycle: Keeping People in Mind throughout Product Design. Elsevier Inc., Amsterdam (2006)
7. Garreta-Domingo, M., Almirall Hill, M., Mor, E.: User-centered design gymkhana. In: Proc. Abstracts CHI 2007, pp. 1741-1746. ACM Press, New York (2007)

Streamlining User Experience Design and Development: Roles, Tasks and Workflow of Applying Rich Application Technologies

Xianjun Sam Zheng, Mo Wang, Gilberto Matos, and Shaopeng Zhang

Siemens Corporate Research, Princeton, New Jersey, USA
{sam.zheng,mo.wang.ext,gilberto.matos,
shaopeng.zhang.ext}@siemens.com

Abstract. The adoption of Rich Application Technologies (RATs), such as Windows Presentation Foundation (WPF) or Adobe Flex, not only significantly enriches the user interface (UI) technology, but also can boost the collaborations among user experience (UX) specialists, designers and developers. Many books and plenty of online resources have described and discussed the technical capability and details of various RATs. However, how to effectively incorporate RATs into the process of UX design and development has not been systematically addressed. In this paper, we report our experience of applying RATs to develop several complex enterprise software systems. A new role, integrator, is introduced to support the communications among UX specialists, designers, and developers. We discuss the responsibilities and task assignments for each role, and propose a new workflow to streamline the design and development. We also discuss the challenges and the lessons learned from applying different RATs.

Keywords: User experience, UI design, development, rich application technology, WPF/Silverlight, Flex.

1 Introduction

One major challenge of delivering successful user experience (UX) is "what you design is not what you will get". During implementation, developers often fall short of reaching the desired design quality defined by UX specialists or created by designers. The reasons can be attributed to the limitations of traditional user interface (UI) technologies, but also to the unidirectional workflow among UX specialists, designers and developers. The introduction and adoption of rich application technologies (RATs), such as Microsoft Windows Presentation Foundation (WPF)/Sliverlight, Adobe Flex, Java FX, not only significantly enrich the UI development technology, but also can boost the collaborations among UX specialists, designers and developers.

Many books and plenty of online resources have described and discussed the technical capability and details of these technologies. However, how to effectively incorporate RATs into the process of user experience (UX) design and development

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 142–151, 2011.
© Springer-Verlag Berlin Heidelberg 2011

has not been systematically addressed. In this paper, we report our experience of applying RATs to develop several complex enterprise software systems. A new role, integrator, is introduced to support the communications among UX specialists, designers, and developers. We discuss the responsibilities and task assignments for each role, and propose a new workflow to streamline the design and development. We also discuss the challenges and the lessons learned from applying different RATs.

1.1 Waterfall Process of Software Development

Traditionally, software design and development follow a waterfall process (e.g, Cusumano & Selby, 1997; Royce, 1970). Figure 1 shows a simple example of this process. Based on the analysis of user needs and requirements, UX specialists create sketches or wireframes to define the system layout and also the interactions among different components. Designers will work on the visual design or "look & feel", adding visual style and drawing necessary icons. These "static designs" or mock-ups, often in the format of JPG, PNG, or BMP, will be delivered to developers, who will program business logic and behaviors to make the UIs functional. Custom or nonstandard UI controls, even though they can be an integral part of novel user experience, are often dropped out or changed during development because it is too difficult or requires too much effort to implement using complex technologies, such as graphics device interface (GDI+). Furthermore, because the whole workflow is unidirectional or one-way, it is rather time and budget consuming if the design needs to be changed at the end of development cycle.

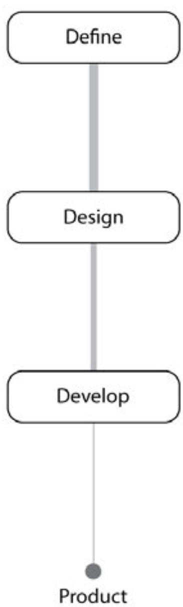


Fig. 1. An illustration of the traditional, one-way workflow for design and development

1.2 Model/View/Controller (MVC) Pattern

To support the concurrent design and development, the ideal process is to decouple presentation and user interaction from business logic and data access following the Model/View/Controller (MVC) pattern. Under MVC, the view manages the presentation or visual feedback to the user; the controller interprets the input or events from the user, such as mouse clicking or keyboard pressing, and manages the model information or updates the view, the model manages the data, responds to requests for information about its state from the view, and responds to instructions to change state from the controller (e.g., Krasner & Pope, 1988).

MVC pattern has been successfully applied in the web application development (e.g., Leff & Rayfield, 2001), where the view is the look of a HTML page, and the controller is the code that collects and generates dynamic data, and model is represented by the content, which is often stored in a database or XML files.

Several rich application technologies, such as WPF, Flex, Java FX, provide implementation support that leverages the MVC pattern. Two most representative examples are WPF and Flex, and both of them have mature and optimized tool support, which plays an important role in their successful adoption by the developer community.

2 Rich Application Technologies

2.1 Windows Presentation Foundation (WPF)

WPF, introduced in the Microsoft .NET Framework 3.0, features a new programming paradigm as an alternative to the traditional Windows Forms programming (e.g., Nathan, 2006; Petzold, 2006). WPF effectively permits the separation of UI from the underlying business logic using a new declarative language: Extensible Application Markup Language (XAML), which is based on XML. The creation of XAML intends to facilitate the communication and collaboration between designers and developers. For instance, detailed designs can be specified in XAML by the designer, and directly used by the developer in implementing the system. The development process can also start from the developer, who first creates components using default styles, then delivers these styles in XAML form to the designer, who then can style it and can completely redesign its appearance.

In WPF, styles define the appearance and interactive behaviors of the elements in an application. A resource is defined for re-use purpose. The external resource dictionary files allow designer and developers to re-use the same asset across multiple applications, providing a consistent user experience for different applications in a suite. Another unique capability of WPF is that similar code base can be used to produce stand-alone desktop application and browser application. This feature can dramatically reduce the efforts and costs for maintaining two different code bases for stand-alone or web-based deployments.

In addition, Microsoft also introduced the Expression Studio, including tools like Design and Blend. The idea is that designers will be using Expression Design, which supports vector based graphics. All the graphics and design generated by designers can be stored using the XAML format, and referenced from both the XAML GUI specification, and from underlying implementation code. Several implementation languages, including C# and Visual Basic .NET typically provide the business logic implementation, whereas XAML delivers specification of the presentation layer. In the case of more advanced visual customizations, the programming languages can be used to explicitely manipulate the styles of visual elements, but even in these situations, the XAML based styles are often reused and dynamically associated with the control elements. Developers can use both Blend and Visual Studio to code XAML and C#.

Blend is an important tool of Expression Studio not only because it has comprehensive functionalities, but also because it has a full span of features in the product development. Starting from Blend 3, a new project type, SketchFlow, has been introduced into both WPF and Silverlight. SketchFlow enables UX specialists to visually define the UI architecture and navigation of the whole system. It also provides the low-fidelity theme for controls. The deliverable of SketchFlow is an application embedded in a navigation panel with notation and commenting tools for other people, such as stakeholders, to provide feedback. Blend can also generate a Word document template embedded with the design. This can also be used to facilitate the communications and discussions of the design details and specifications among team members.

2.2 Adobe Flex

Flex is a SDK (Software Development Kit) created by Adobe company for building highly interactive applications that can be deployed on Internet browsers installed with Flash Player plug-in or desktops installed with Adobe AIR. The UI layout and graphical appearance are described by MXML, a declarative XML-based language. The programming language used to define business logic is ActionScript which is an object-oriented language. Flex has included a rich library of basic UI components, such as buttons, list box etc., as well as some advanced UI component, such as data grid, data graph etc. Flex framework has its advantage in flexibility. For instance, Flex at its core is event-driven and has many mechanisms to incorporate MXML and ActionScript to implement complex behaviors. This flexibility makes it easy for developers to quickly build prototypes.

Flash Builder (formerly called Flex Builder) is an Eclipse-based IDE (integrated development environment) for rich internet applications (RIAs) and cross-platform desktop applications development with the Flex framework embedded in. Its main features include intelligent coding, interactive step-through debugging, and a WYSIWYG editor for the user interface layout and appearance design.

Another tool, Flash Catalyst, has been introduced in 2010, which aims to rapidly create expressive interfaces and interactive content. As an interaction design application, Flash Catalyst intends to be a bridge that can bring UX specialists, designers, and developers together. The tool is fully incorporated in the Adobe design suite and can import graphic assets created in Adobe Photoshop and Illustrator. It also can export a skeleton Flex/AIR project to Flash Builder, allowing developers to implement the business logic using Flex framework. However, the primary usage of Flash Catalyst is to create wireframes and interactive projects without writing code. For instance, it does not even support code editing, nor does it support two-way workflow with Flash Builder. It does, however, support the two-way workflow with Illustrator and Photoshop. This allows designers continue working in Illustrator or Photoshop to refine design and their art work, which later can be imported into the wireframes in Flash Catalyst. We view Flash Catalyst as a tool primarily for UX specialists and partially for designers. Although it is a tool with limited functionalities and targeting at specific set of users, Flash Catalyst is easy to learn and used, especially suitable for people with less technical background.

3 Roles, Tasks and Workflow

In the past several years, we have applied both WPF and Flex to develop several complex enterprise software systems across various domains, such as building management, healthcare IT, and energy services. RATs were chosen as the major front-end development technology for a couple of reasons. First is due to their flexible deployment model to either desktop or web browser. Second, the separation between UI presentation and functionality is suitable for the collaboration among different teams at different locations and time zones. Lastly, many new UI features offered by both WPF and Flex, such as animation, customer control, and data visualization, empower the designer to create a new and improved user experience.

Because of the large scope of the projects, multiple teams with different expertise, including UX specialists, designers, developers, were involved in the development. A new role, integrator, was introduced to support the communications among designers and developers. We first discuss the tasks and responsibilities for each role, as summarized in Table 1. We also discuss the workflow for the design and development, which is illustrated in Figure 2.

3.1 Different Roles

UX Specialist. The primary responsibility for UX specialists is to define and ensure product or system user experience. The tasks include 1) understanding and analyzing end-user's requirements, such as their needs, capabilities, and behaviors in their real-world environments; 2) then translating the analysis results of users' tasks, use cases, and workflow into UI layout and interaction models. The deliverables are wireframes, sketches, and storyboards. The typical tools UX specialists used are PowerPoint, Visio, etc, which are independent to any RATs. However, both SketchFlow and Flash Catalyst can be useful tools for UX specialists to create wireframes, storyboards, and some quick prototypes. Based on our experience, Flash Catalyst seems to be easier to learn, but provides limited tool support for collecting user feedback.

Designer. The main responsibility for designers is to design "look & feel" based on the specifications defined by the UX specialist. Designers are used to create static screens, mockups, and icons using primarily Adobe design tools, such as Photoshop or Illustrator. The introduction of RATs can have significant impact on designer's tasks. For instance, with WPF, designers no longer work on mockups or screens but the actual XAML code (even though Expression Design allows designers to produce assets like they do with Illustrator or Photoshop and then deliver them in XAML, the functionality for design is quite limited). The deliverables are XAML assets, including icons, brushes, animations, appearance/skin, shape, color, and resources. Designers would mainly use Expression Design and Blend tools, which allow them to actually execute the implemented functional systems, and then to directly manipulate the visual resources and styles. The main benefit is that designers can have full control of their design, removing the need for a roundtrip communication and verification of the changes made by the developers. However, because both Expression Design and Blend are new to designers, some training for the tools is necessary. In fact, based on our experience, designers, especially those who don't have programming background, will have a steep learning curve for the Blend tool.

With Adobe Flex, the change for designers is minimal because designers are still using the same design tools, such as Photoshop or Illustrator. The main difference is that their deliverables are files in FXG format (Flash XML Graphics), which are XML graphics files, can be imported directly into Flash Catalyst or later into Flash Builder.

Developer. The primary responsibility for developers is to develop functionality. Their tasks include, 1) understanding and realizing the UI implementation architecture that supports the required user interaction patterns; 2) leveraging the existing reusable assets and the development team expertise toward a full functional implementation; and 3) building connections to the data source. They create a functional UI according to the specifications defined by UX specialist and integrator, but without spending extra time on customizing the look and feel. They implement interactive functionality, bind the UI elements to the corresponding data source, and implement event handlers triggered by the user inputs and actions. The deliverables for developers include skinless UI controls and data binding. In WPF, they use primarily Visual Studio tool and the programming language is .NET programming languages such as C# and VB.NET. Whereas, in Adobe Flex, they use Flash Builder and the programming language is ActionScript.

Integrator. The integrator is a new role, who should have extensive training in programming and also a good sensibility to design. In the early design phases, the integrator needs to inform the UX specialists and designers of the specific constraints inherent in the selected implementation technology (e.g., WPF or Flex) that they should be aware of. This type of interaction continues throughout the life of the project, with the adoption of additional architectural decisions and implementation constraints being discovered. His/her responsibility is to assign tasks (functionality implementation or skin design) to developers or designers. Finally, the integrator is heavily involved in the production of a complete system, including the designers' declaration of visual artifacts, and the functionality and behavior implementation of the UI, coded by the developer. In WPF, the major tool used by the integrator include is Expression Blend. In Flex, the integrator will be mainly using Flash Builder and Flash Catalyst.

3.2 Workflow

RATs support concurrent design and development, and the integrator works on synchronizing the concurrent specialized teams. Consequently, a new workflow is proposed to streamline the UX design and development as illustrated in Figure 1.

Table 1. A summary of different user roles and tasks, tools for WPF (rows with white background) and Flex (rows with light gray background)

Roles	Tasks	RATs	Language	Tools
UX Specialist	Define UI concept, create wireframes & sketches	WPF		PowerPoint, Visio, SketchFlow
		Flex		PowerPoint, Visio, Flash Catalyst
Designer	Design visual “look & feel”	WPF	XAML, styles & resources	Expression Design, Blend
		Flex	MXML, graphic assets	Adobe Illustrator, Photoshop
Developer	Implement functionalities & interactive behaviors	WPF	C#, VB.Net, XAML components	Visual Studio
		Flex	MXML/ActionScript	Flash Builder
Integrator	Define UI architecture, coordinate and integrate the activities of UX specialist, designer, and developer	WPF	XAML/C#	Expression Blend
		Flex	MXML/ActionScript	Flash Builder

UX specialists, designers, and developers no longer work in a waterfall or one-way workflow. In fact, defining, designing and developing can happen concurrently, and there are a lot of two-ways communications between UX specialists and designers, UX specialists and developers, and also designers and developers. However, these communication and collaboration need to be coordinated by the integrator, who will sometimes work closely with UX specialists to understand and implement the UI architecture (e.g., screen layout), sometimes work closely with designers to provide technical support to generate XAML or MXML code, and often work closely with developers to make sure functionality implementation can be seamlessly integrated with the design.

Note that both UX specialists and integrators are responsible for the UI quality, with UX specialists focusing on the design while integrators focus on the realization and implementation. Both UX specialist and integrator are talking to designers and UI developers; but the UX specialist has more influences on designers than the integrator; conversely, the integrator has more influence on the UI developers. The integrator is also responsible for the communications between the backend and the frontend development. Sometime, there are overlaps between designers and developers, especially for the UI control interaction and animation. Designers can directly create some animations using XAML or MXML; but for special animations, it is much easier to implement in C# or ActionScript by the developer. The integrator needs to anticipate these issues, and make decisions accordingly.

This new workflow, on the one hand, encourages the communication and collaborations for different team members with different expertise, on the other hand, ensures that the efforts from different team members, such as UX specialists, designers, and developers, can be integrated together to deliver the desired design quality. As the result, it can achieve the outcome of "what you design is what you get".

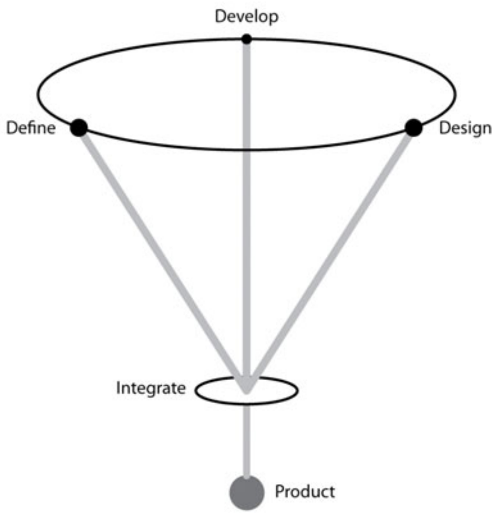


Fig. 2. An illustration of the new workflow for UX design & development

4 Lessons Learned and Discussions

Overall, RATs and their associated tools offer significant support for concurrent software development with multiple teams with different expertise. Our experience shows that with clearly defined roles and task assignment, as well as a new workflow, we can utilize the power of RATs to streamline the UX design and development. Nonetheless, there will still be significant challenges in developing advanced interactive system, even when fully leveraging the capabilities of RATs.

4.1 Non Technical Lessons

First, even though new technology is in place, it takes time for developers and designers to adopt the new process and development model. For instance, in WPF, instead of delivering static graphics or screen that they used to do, designer should deliver production artifacts, preferably in XAML format as a resource dictionary that can readily be used by the controls used in the implemented application. Developers should only focus on the functionality, and should pay no attention to the "look" of the UI, only to using the right controls that will use the proper visual skin from the resource definition. Secondly, the integration between designers and developers needs to be planned ahead rather than ad-hoc. Namely, the integrator should work closely with UX specialist to lay down the UI architecture, and define the UI controls. Based on the analysis, the integrator shall decide whether a control template or data template is needed. Thirdly, the learning curve of WPF can be steep, especially for designers who have no programming training or experience.

4.2 Technical Lessons

Data binding. Both WPF and Flex generally create a good framework for separation of concerns between the UI designers and application developers, and allows the two groups to develop their artifacts independently, with or without tight integration of changes. This decoupling tends to break when the UI design tries to push the limit of the dynamic visual controls, where the controls need to be tightly coupled to the data bound objects for some aspects of graphical definition. In these situations, the data binding mechanism becomes a glue technology for the implementation of the appropriate object structures, and their mapping to the dynamically controlled interactive controls. For integrators, Blend provides more powerful tools for data binding to generate sample data of common types (e.g. name, address, email, phone number). In Flash Catalyst, the integrator would have to manually input the sample data.

Style (reuse control group). WPF and Flex provide a structured resource management framework in Expression Blend and Flash Builder, allowing the UI designers to create and replicate visual styles in resource dictionaries. While these capabilities enable the quick initial creation of visual resources, their effective maintenance becomes very time consuming if the stylistic dependencies are not managed by reusing shared aspects of the design elements. In general, the definition and maintenance of the shared and reused resource structures is the type of task that fits better with the software developers than with the UI designers' skillset, but it still requires a keen understanding of the visual dependencies as well. This aspect of style and resource management should be handled or overseen by the integrator who needs to have the knowledge of both the visual and structural aspects of the application architecture, as well as the planned or potential changes.

Non-standard behavior. Many components in WPF and Flex provide a standard behavior and allow wide customizations, both in their visual presentation and interactive behavior, as do a multitude of external component libraries. The customization capability often tempts UX specialists to try to define optimized interaction and behavior capabilities which provide a marginal benefit to the application user over the standard behavior. Unfortunately, changes in one aspect of behavior for complex user controls may require extensive rework and impact many other aspects due to the implementation dependencies. The actual cost and effort for such customizations in many cases may make them non-cost effective, particularly if the expected benefit is small. In addition, once a control is extensively customized and modified, it often needs to be maintained to preserve the customization with new releases of the base components. In general, only if a major competitive advantage of a system depends on modifying the standard control containers, can such a decision be justified.

5 Conclusion

The emerging technologies, RATs, are making it easier for UX specialists, designers, and developers to work collaboratively on developing new applications, by concentrating on their respective specialties and having a well defined integration interface. This interface is standardized on the technology level and does not map to the domain specifics of any application, so there is still much left to interpretation of the assets that are coupled over the interface, and to the common interface understanding between designers and UI developers. The management of this communication interface, which is both complex and project critical becomes the primary responsibility of the integrator that collaborates with the designers and developers in making sure the shared resources are defined and used in accordance with a common interpretation of the design and behavior requirements.

Acknowledgments. We thank Laurent Bugnion for valuable discussions and inputs. We thank Miao Wang & Nicola Vittori for helping create the visual illustrations.

References

1. Cusumano, M., Selby, R.: How Microsoft builds software. Communications of the ACM 40(6), 53–62 (1997)
2. Leff, A., Rayfield, J.: Web-App. Dev Using the M/V/C Design Pattern. In: Fifth IEEE Intl. Enterprise Distributed Object Comp. Conference (2001)
3. Krasner, G., Pope, S.: A cookbook for using the model-view controller user interface paradigm in Smalltalk-80. Journal of Object-Oriented Programming 1, 26–49 (1988)
4. Nathan, A.: Windows Presentation Foundation Unleashed (WPF). Sams Press, USA (2006)
5. Petzold, C.: Applications= Code+ Markup: A Guide to the Microsoft Windows Presentation. Microsoft Press, Redmond (2006)
6. Royce, W.: Managing the Development of Large Software Systems. In: Proceedings of IEEE WESCON, vol. 26, pp. 1–9 (August 1970)

Part III

Model-Based and Patterns-Based Design and Development

Configurable Executable Task Models Supporting the Transition from Design Time to Runtime

Birgit Bomsdorf*, Stefan Grau, Martin Hudasch, and Jan-Torsten Milde*

Department of Applied Computer Science, University of Applied Sciences Fulda, Marquardstr. 35, D-36039 Fulda, Germany
{bomsdorf,milde}@hs-fulda.de

Abstract. Model-based design of user interfaces mostly starts with task and domain modeling. The resulting models are an important input to subsequent development steps. Thus, a thorough evaluation of these specifications is of great importance, e.g. to avoid the implementation of bad or even error prone solutions. Executable task models are in use for several years to evaluate the design time specifications. They are also used at runtime by now as part of the final application. In this paper we propose an executable task model that is configured using the design time model. Kernel concept of this work is a task state machine describing a generic task life cycle assigned to each task. Developers may extend it at design time to describe application dependent behavior. The extensions are automatically transferred to the runtime system. A further focus of the paper is on the specification of temporal relations and their extensibility in terms of model description and execution.

Keywords: task modeling, executable task model, runtime support, temporal relations.

1 Introduction

In model-based design of user interfaces it is well established to distinguish different modeling layers [1]: Starting with a model of the tasks formulated from the users' perspective (*task model*) and a model of entities involved in task execution (*domain model*) a model of the abstract user interface (*dialog* and *presentation model*) is developed. This model is transformed into one or more concrete UI models taking into account context information such as availability of media, modes and devices. All in all, results of task modeling are an important input to the subsequent development steps. A task model, however, may be very complex. It describes not only the tasks users should perform to reach their goals by means of a system under consideration, but also their hierarchical decomposition into subtasks, information on task sequencing and conditions of task execution. . Executable task models are in use for several years to evaluate such specification [2], [3], [4], [5]. They are also used at runtime by now as a component of the final application [6], [7], [8], [9], [10].

Executable task models at runtime are based on generic task life cycles in the works mentioned above. They are represented by means of state machines showing a

* Corresponding authors.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 155–164, 2011.
© Springer-Verlag Berlin Heidelberg 2011

lot of similarities with respect to states and transitions. In our work, in contrast to the other approaches, the state machine is not only part of the runtime system but additionally used as an explicit modeling concept at design time [11]. Specifications by the developers extend the generic state machine of a task and are systematically adopted by the runtime system. Basically, similar to [8] and to [9], we follow the approach of model interpretation at runtime. The design time models are used to configure the executable task model. This work aims at supporting the transition from design time to runtime. Its main ideas are presented in the following. In section 2 basic modeling concepts including the generic state machine are introduced. In the subsequent section 3 the generic state machine as used at runtime and steps of the configuration process are presented. Remarks concerning the state of the work are given in section 4.

2 Modeling Concepts

The kernel modeling concepts of the approach presented here are *Task*, *Condition*, *Temporal* R*elation*, *Behavior* and *Domain Object* that are linked to each other via different, predefined relations. The concepts are shortly introduced below except *Domain Object*, which is used to specify objects that are involved in task execution. The object model is not further described here since in our current work we focus on tasks and temporal relations. Basically an object-oriented approach is used that could be described by means of UML. Examples of linking objects with other concepts are given within the following introduction of the concepts *Behavior* and *Condition*.

2.1 Task Specification

Task structures, as in general, are defined by means of hierarchical decomposition of tasks (*CompositeTask*) into their subtasks. A task not subdivided into further tasks is called an elementary task (*ElementaryTask*). These bottom tasks of a task hierarchy are the "points" where the intended actions affecting the domain are performed. Composite tasks are high level tasks structuring a task space.

The time period within which a task is executed is used as an explicit concept to describe temporal dependencies. A task defines a frame (*task frame*) of its execution and thus for the actions performed to fulfill it. In the case of a composite task each subtask possesses a frame of its own while the superior task's frame includes all of them. Overall, temporal dependencies are based on time periods (as in the work by Allen [12]), e.g. constraints on task duration (temporal constraints), but also on temporal relations (as often used in task modeling, e.g. in MAD [5] and CTT [3]).

Temporal Constraints. *Temporal constraints*, defined as task attributes in the meta-model, specify delays before (*StartOffset*) and after (*EndOffset*) a task run. In addition, the duration of task execution can be constrained in terms of a minimum and a maximum period of time. For example, a user might move the icon of a text document into a workspace by means of a defined touch-based gesture. The user can repeat this several times but the system should take care of a clear separation of the different repetitions. Fig. 1 (a) shows a solution: The required behavior is modeled by a *Loop* of the interaction task *move icon into workspace* together with a value specification for its *EndOffset* defining a delay between two executions. *MinDuration* and *MaxDuration* might be set to specify the period of time within which the gesture is to be performed.

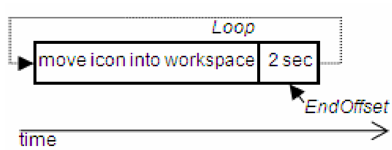
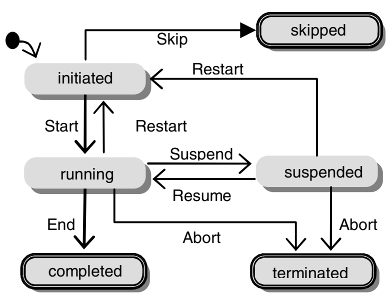


Fig. 1. Loop of a task with a time interval of 2 seconds after each performance

Generic Task Life Cycle at Design Time. Tasks undergo different state changes while being performed. These states are also significant to users since in their planning of follow-up activities they take into account current task situations. It is important to a user, for example, whether he can start to work on a task or not (because of unfulfilled conditions), or if he is already performing a task and thus its subtasks. The states and possible state changes are described by means of a generic state machine. It and the description of its states are shown in Fig. 2. The transitions between the task states are labeled with the events triggering them.



State	Meaning
initiated	if all preconditions are fulfilled the task can be started
skipped	the task is omitted
running	denotes the actual performance of the task and of its subtasks, if applicable
completed	marks a successful task execution
suspended	the task is interrupted
terminated	indicates an abort

Fig. 2. Generic task state machine used at design time [11]

Developers may extend each transition at design time, e.g., to denote that completing a task *logout* triggers the abortion of a task *check out*. In this example, the transition from *running* to *completed* of the *logout* task state machine may be extended by the generation of an *Abort* event that is sent to *check out* and its task state machine, respectively. Further actions may be added to transitions, e.g., to specify modifications of values of domain objects or to alter the Boolean values of constraints [11]. In addition to these effects resulting from task state changes the intended effect of a regular task execution is described by a so-called *Behavior*.

2.2 Behaviors and Conditions

Behaviors specify the effects of task execution, i.e. the intended actions affecting the domain. They can be assigned only to elementary tasks (performed while the task is in the state *running*). Thus, a composite task doesn't possess any behavior. Its effect is given by the combination of its subtasks, which themselves may be elementary or composite tasks.

The meta-model contains different specializations of the concept *Behavior*, e.g. *ObjectManipulationBehavior* by which manipulations of domain objects are specified. Developers may extend the set of behaviors as needed. All in all, our way of connecting specific manipulations to task execution is similar to the use of *userAction* and *taskItem* (object of an action) described in [13, page 62]. Currently we also follow an object-oriented approach, thus an *ObjectManipulationBehavior* contains a field referencing an object's method. In the case the behavior is assigned to a task, the method is invoked if the task is performed. A further example of a behavior type is the *RequestTaskStateChangeBehavior* by which impacts on other tasks performance may be specified. Such a behavior is used in the example below that is given by showing the XML description internally used. If the *AbortCheckOut* behavior is executed an *Abort* event is sent to the task with the identification *46…* (line 3 to 5).

```
1 <Behaviours>
2 <RequestTaskStateChangeBehaviour global:Id="22…"
global:Name="AbortCheckOut">
3     <TaskInstanceReference Reference="46…"/>
4     <RequestedTaskState>Abort</RequestedTaskState>
5   </RequestTaskStateChangeBehaviour>
6  </Behaviours>
```

Conditions constrain task performance, e.g. a task can be started only if its pre-conditions hold true, while it can be finalized only if its post-conditions are fulfilled. The concept *Condition* enables to cope with such constraints. Similarly to behaviors predefined types exist, for example, an *ObjectValueCondition* is true if an attribute of a domain object holds a given value, otherwise the condition is false. By means of a *TaskHasStateCondition* it is evaluated whether or not a task is in a given state. In the example below *checkout is performed* is a *TaskHasStateCondition*. It is true if the task with the ID *4D…* is in its *Running* state.

```
<Conditions>
<TaskHasStateCondition global:Id="30… " global:Name="checkout is
performed" …>
    <TaskInstanceReference Reference="4D…"/>
    <TaskState>Running</TaskState>
  </TaskHasStateCondition>
 </Conditions>
```

The following description exemplifies how behaviors and conditions are assigned to tasks by means of references. The condition *checkout is performed* is assigned as a pre-condition to the task *empty shopping cart* (line 2 to 4). The effect of this task is given by the behavior *AbortCheckOut* (line 6). Generally, usage of a condition as a pre- or post-condition depends on whether it is referenced in the *PreConditions* or in the *PostConditions* section.

```
1 <ElementaryTask global:Id="55…" global:Name="empty shopping cart"
Type="Interaction">
2 <PreConditions>
3    <ConditionReference Reference="30…"/>
4  </PreConditions>
5  <PostConditions/>
6  <BehaviourReference Reference="22…"/>
7 </ElementaryTask>
```

2.3 Temporal Relationships

A *TemporalRelationship* characterizes task sequencing and hereby the initial use of a set of modalities over time. A set of predefined relationships exists, such as *Sequence* (sequential task execution), *Parallel* (timely overlapping of task execution), *Loop* (repeated task execution). Further patterns, however, can be specified and added to the set. Temporal relationships are used within so-called temporal relationship structures to model the order of subtask execution.

A temporal relationship structure is described by means of a hierarchy within which a temporal relationship may include further ones, i.e. temporal relationships, similar to tasks, are structured hierarchically. Inner nodes represent tasks or further temporal relationships, but a leaf node must be a task. Hereby complex temporal dependencies can be specified - without introducing additional tasks only for the purpose of defining temporal ordering as it is required by, e.g., MAD [5] and in our former approach WTM [11].

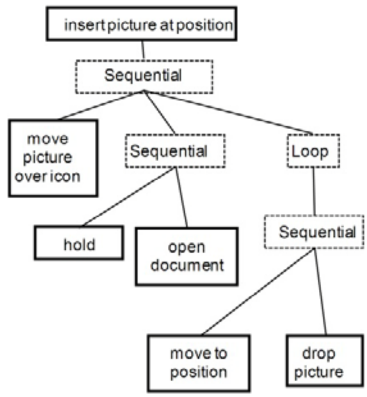
In a multimodal interface for example, the user may gaze at a text document icon while repeatedly moving a picture causing the picture to be inserted at the beginning of the text document. In this example (see Fig. 3 (a)) the task *insert picture* exists that is composed of the elementary subtasks *gaze at icon* (possessing the ID *71…*) and *move picture* (ID *5E…*). Subtasks are listed within the *CompositeTask* TAG (line 3 and 4) while their execution order is given by means of the *TemporalRelationships* Tag (line 6 to 13).

If the user, however, holds the picture over the icon for more than two seconds the document should be opened. Now the user may specify an exact position by moving the picture over the opened document and dropping it once the desired position is reached. Fig. 3 right hand shows the corresponding temporal structure. This time a tree notation depicting the temporal hierarchy instead of XML specification is used. It shows in terms of the example how a hierarchical temporal relationship structure is composed of subtasks and temporal relationships (rectangles with dotted lines).

Generally, to each composite task a temporal relation structure is assigned. The occurrence of each subtask within the tree is restricted to exactly one: The temporal relation structure is defined for all subtasks the superior task is composed of, but each subtask can be referenced only one time.

```
<CompositeTask … global:Name="insert picture"…>
    …
            <ElementaryTaskInstance Id="5E..." …/>
            <ElementaryTaskInstance Id="71..." …/>
    …
    <TemporalRelationships>
        <Parallel>
            <Loop>
                <TaskReference Reference="5E…"/>
            </Loop >
            <TaskReference Reference="71…"/>
        </Parallel>
    </TemporalRelationships>
</CompositeTask>
```

(a) a temporal structure given by means of XML specification



(b) a temporal structure visualized by hierarchical tree notation

Fig. 3. Temporal structures of inserting a picture

3 Executable Task Model

The executable task model is part of a runtime environment that is configured based on the static model specified at design time. The environment contains a *TaskManager*, a *DomainManager*, a *BehaviorManager*, a *ConditionManager*, and a *TemporalRelationManager* that interact which each other to realize the behavior as given by the static model. The managers responsible for behaviors, conditions and temporal relations are extended with the implementation of specific behaviors (such as *ObjectManipulationBehavior* and *RequestTaskStateChangeBehavior*), conditions (e.g. *ObjectValueCondition* and *TaskHasStateCondition*) and temporal relations (*Sequence*, *Parallel*, *Loop*, etc.), respectively. Adding, for example, a further temporal relation to our modeling approach requires two steps: First of all, the meta-model, i.e. its XML schema has to be extended with an according Tag so that the additional temporal relation can be used at design time. Secondly, the relation must be coded and added to the runtime environment.

3.1 Generic Task Life Cycle at Runtime

The kernel concept of the executable task model is a generic task lifecycle to handle and control task performance. It means that the approach presented in this paper makes use of a generic task lifecycle not only at design time (design task state machine, in short d-TSM) but also at runtime (referred to as r-TSM).

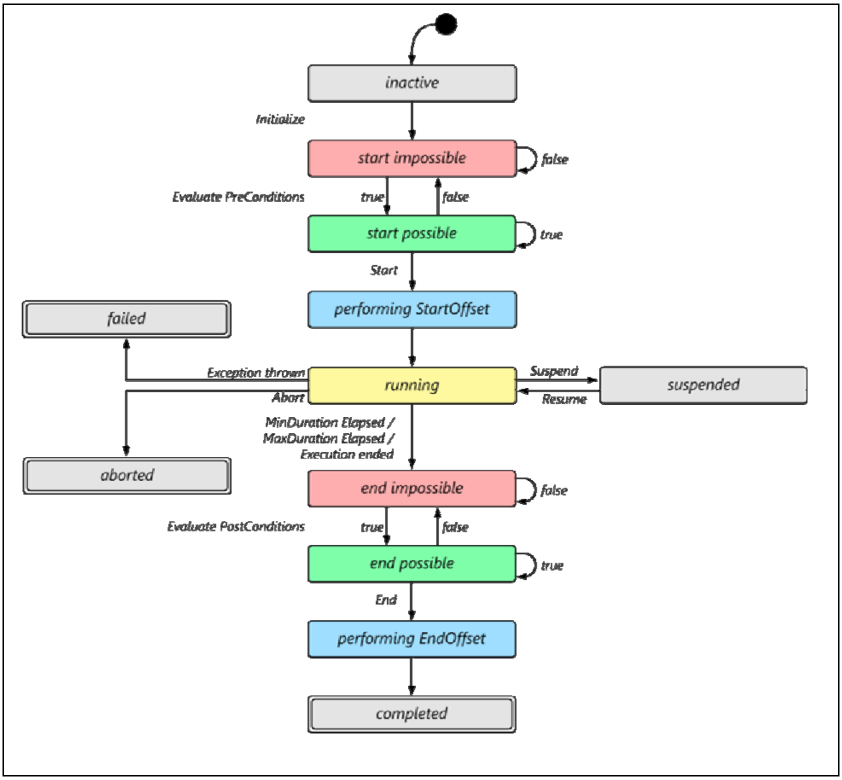


Fig. 4. Generic task state machine used at runtime

The generic task state machine of the runtime environment (see Fig. 4) is an extended version of the one introduced in Fig. 2 since additional runtime control information is required to handle task execution. All in all, the d-TSM describes the task life cycle from the user's point of view while the r-TSM describes it from the system's point of view. (Please note that the skip possibility is handled by the *TemporalRelationManager* and thus not modeled explicitly in the runtime task state machine.) The meanings of the states are as follows:

inactive: The task is part of the task space but not "accessible", i.e. the software object representing the task exists but is not initialized.

start impossible: Beginning the task is not possible because at least one of its pre-conditions is not fulfilled. Conditions are checked permanently. In the case a condition changed its Boolean value the task manager is informed, so that for all transitions depending on that condition it is evaluated whether they are to be triggered. A task remains in the state *start impossible* as long as at least one pre-condition is unfulfilled (false) and the state *start possible* is adopted as soon as all pre-conditions are true.

start possible: All pre-conditions of the task are fulfilled. The task state is set to *start impossible* again once one of the pre-conditions evaluates to false. However, if the

Start event is received the task begins to be performed. This event may result from a user action and thus be generated by the UI.

performing StartOffset: The task is started and a defined time interval is waited before it actually runs. As soon as the *StartOffset* elapses the state *running* is adopted. If no *StartOffset* is defined (i.e. its value is zero), the state *running* immediately becomes the current state.

running: This state denotes the actual performance of a task and it's subtasks, if applicable.

suspended: The task is interrupted because of the application logic, i.e. the events *Suspend* and *Resume* are triggered by the application.

end impossible: The task run cannot be ended because at least one of its post-conditions is not fulfilled.

end possible: All post-conditions of the task are fulfilled. However, ending the task may depend on further dependencies. Therefore, only the *End* event denotes the end of task execution. Again, the event may result from UI interactions.

performing EndOffset: The task is ended and a defined time interval is waited before it is completed. If the value of the *EndOffset* is zero, *end possible* immediately becomes the current state.

completed: This state marks a successful task execution.

aborted: This state indicates an abort that is triggered by the application logic.

failed: The task performance fails because an exception is thrown.

3.2 Configuration

A state machine exists for each task in both the design model and the executable task model. Extensions of a d-TSM are transferred systematically to the r-TSM within the configuration process. Each transition extension of a task's generic d-TSM is added to the respective transition of its generic r-TSM. Hence, if the *TaskManager* performs a state transition all actions assigned to it are invoked. For example, in the case a transition is extended with a *RequestTaskStateChangeBehavior* a state-change event according to the given state in the behavior specification is sent to the task referenced in the behavior specification.

Conditions are introduced into r-TSM according to their specified usage in the design model: Pre-conditions of a task are assigned to its transitions between *start impossible* and *start possible* as (additional) requirements for performing these transitions. Post-conditions, in the same way, extend the transitions between *end impossible* and *end possible*.

StartOffset and *EndOffset* are represented by means of the states *performing StartOffset* and *performing EndOffset*, respectively. As described by the generic r-TSM, the specified time has to elapse before the subsequent state is adopted. Fulfillment of durations constraining task execution (*MinDuration* and *MaxDuration*) is not realized by an explicit state but controlled by the *TaskManager* while the task is in *running*. Additionally, in the case of a composite task the r-TSM of its subtasks are considered by the execution process according to the defined temporal structure.

Temporal relations are systematically transferred to the executable model as well. Here the hierarchies of both the tasks and the temporal structures are regarded. Thus, temporal relations are fulfilled in close cooperation between the *TaskManager* and the

TemporalRalationManager. A composite task, as mentioned above, possesses a list of all its subtasks and a hierarchical temporal structure. The last one is composed of temporal relations (inner nodes) and task references (leaf nodes). Loading the XML specification of a temporal structure results first of all in the internal representation of its hierarchy. Fig. 5 (a) shows the temporal hierarchy resulting from the example depicted in Fig. 3 (a). The *TemporalRalationManager* instantiates for each relation an according control object. However, this is only possible if the runtime environment contains an extension implementing the relation's semantics. In the case of a temporal relation Tag without corresponding implementation a warning is generated and the relation is ignored by substituting it with the initial *NoOrder* relation. The task references given in the model specification remain but are replaced at runtime by the software objects representing the referenced tasks (see Fig. 5 (b)). Furthermore, the software objects realizing behaviors are analogously linked to tasks. Fig. 6 (b) shows the order of task execution for the simple example *insert picture*, in which the behaviors of *move picture* and of *gaze at icon* are given by *behaviorMP* and *behaviorGI*, respectively.

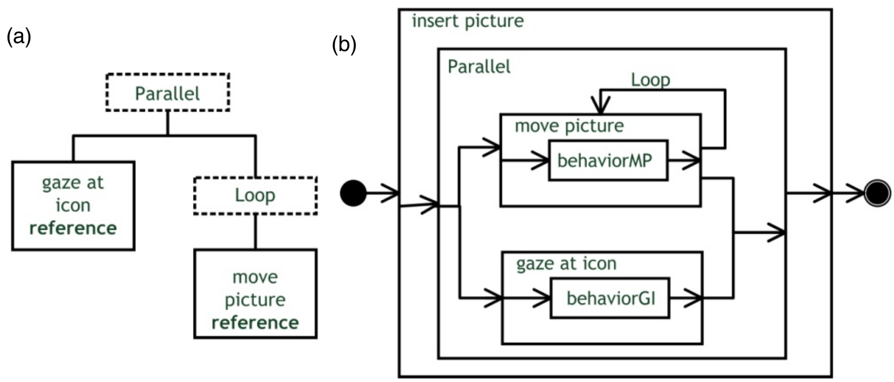


Fig. 5. Temporal Structure after (a) Loading and (b) at Runtime

4 State of Work

The meta-model is formally described by means of XML Schema. First versions of a task editor, a runtime system executing the models and a simulator based on it are developed. Both the editor and the simulation tool are in the state of proof of concept tools. The focus was on extensibility and on temporal dependencies. Easy extensibility was reached by separating the concerns of the concepts *Task*, *Condition*, *Temporal* R*elation*, *Behavior* and *Domain Object* that are linked to each other by means of references, processed by dedicated runtime managers. The focus on temporal dependencies is motivated by multimodal user interfaces, in which complex relations may exists. At task layer we abstract from specific usage of interaction modalities as far as possible and describe the impact of their use on task performance over time. Task frames define the time intervals that are structured by means of hierarchical ordering of temporal relationships and sub-tasks. The editor and the simulation tool mentioned above make use of nearly the same diagrams. The representations and view concepts, however, are to be investigated in follow-up work. A step in this direction is our current work on improving the task state machine editor.

Acknowledgments. This work was supported by the German Federal Ministry of Education and Research.

References

1. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.A.: Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
2. Bomsdorf, B., Szwillus, G.: Early Prototyping based on Executable Task Models. In: Human Factors in Computing Systems, Conference Companion CHI 1996, pp. 254–255 (1996)
3. Mori, G., Paterno, F., Santoro, C.: CTTE: support for developing and analyzing task models for interactive system design. IEEE Transactions on Software Engineering 28(8), 797–813 (2002)
4. Uhr, H.: TOMBOLA: Simulation and User-Specific Presentation of Executable Task Models. In: Proc. of the International HCI Conference, pp. 263–267 (2003)
5. Caffiau, S., Girard, P., Scapin, D., Guittet, L., Sanou, L.: Formally Expressing the Users' Objects World in Task Models. In: England, D., Palanque, P., Vanderdonckt, J., Wild, P.J. (eds.) TAMODIA 2009. LNCS, vol. 5963, pp. 117–130. Springer, Heidelberg (2010)
6. Klug, T., Kangasharju, J.: Executable Task Models. In: TAMODIA 2005 Proceedings of the 4th International Workshop on Task Models and Diagrams, pp. 119–122 (2005)
7. Betermieux, S., Bomsdorf, B.: Task-Driven Composition of Web User Interfaces. In: CADUI 2008, 7th International Conference on Computer-Aided Design of User Interfaces, pp. 233–244 (2008)
8. Clerckx, T., Vandervelpen, C., Coninx, K.: Task-based design and runtime support for multimodal user interface distribution. In: Gulliksen, J., Harning, M.B., van der Veer, G.C., Wesson, J. (eds.) EIS 2007. LNCS, vol. 4940, pp. 89–105. Springer, Heidelberg (2008)
9. Bouchet, J., Nigay, L., Ganille, T.: ICARE: a component-based approach for the design and development of multimodal interfaces. In: CHI Extended Abstracts 2004, pp. 1325–1328 (2004)
10. Blumendorf, M., Lehmann, G., Feuerstack, S., Albayrak, Ş.: Executable Models for Human-Computer Interaction, In Interactive Systems. In: Graham, T.C.N. (ed.) DSV-IS 2008. LNCS, vol. 5136, pp. 238–251. Springer, Heidelberg (2008)
11. Bomsdorf, B.: The WebTaskModel Approach to Web Process Modelling. In: Winckler, M., Johnson, H. (eds.) TAMODIA 2007. LNCS, vol. 4849, pp. 240–253. Springer, Heidelberg (2007)
12. Allen, J.F.: Maintaining Knowledge about Temporal Intervals. Communications of the ACM 26(11), 832–843 (1983)
13. Stanciulescu, A.: A Methodology for Developing Multimodal User Interfaces of Information System, Ph.D. thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium (2008)

Automatic Adaptation of User Workflows within Model-Based User Interface Generation during Runtime on the Example of the SmartMote

Kai Breiner[1], Kai Bizik[1], Thilo Rauch[1], Marc Seissler[2], Gerrit Meixner[3], and Philipp Diebold[1]

[1] Software Engineering Research Group, University of Kaiserslautern, 67663 Kaiserslautern, Germany
[2] Center for Human-Machine-Interaction, University of Kaiserslautern, 67663 Kaiserslautern, Germany
[3] German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany
{Breiner,K_Bizik,T_Rauch,p_diebol}@cs.uni-kl.de, Gerrit.Meixner@dfki.de, Marc.Seissler@mv.uni-kl.de

Abstract. Model-based universal interaction devices are already capable to react on contextual changes by automatically adapting the user interface, but without considering the usefulness of the resulting user interface. Often tasks cannot be executed any more or execution orders will result in dead locks caused by unavailable functionality. We present our approach of investigating this property of adapted models based on the example of the SmartMote in our living lab the *SmartFactory*$^{\mathbf{KL}}$. Given the task description of the user interaction we determine a dialog model in terms of a state machine – which is necessary in our process of user interface generation – to determine possible execution orders leading to the accept state of this state machine. Using these execution orders the initial task model can be adapted, all misleading tasks can be removed and the resulting user interface will only offer valid user interactions.

Keywords: Adaptive User Interfaces, Usage Context, Task Fulfillment, SmartFactory, SmartMote, Model-based User Interface Development.

1 Introduction

Modern production environments comprise numerous devices demanding for direct human interaction – e.g. to execute, monitor, or maintain the production process. While each of this devices confront the user with different user interfaces – providing different user experience – future facilities will be equipped with universal control devices such as the SmartMote (see Figure 1, right) that are capable of providing control over intra-device functionality in a homogeneous manner, enabling for a homogeneous user interaction concept. This can be ensured by the separation of the content (i.e. task/workflow model) and its representation (i.e. presentation model). Due to the flexible nature of future ambient production environments, production lines can be (automatically) rearranged e.g. in order to compensate for device errors.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 165–174, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Thus, devices and their services will not be available anymore which needs to be reflected by the user interface in order to prevent usage errors. Therefore, the underlying model will be tailored to the new configuration. This paper discusses one approach based on the adapted model to determine if the user is still able to achieve her goal.

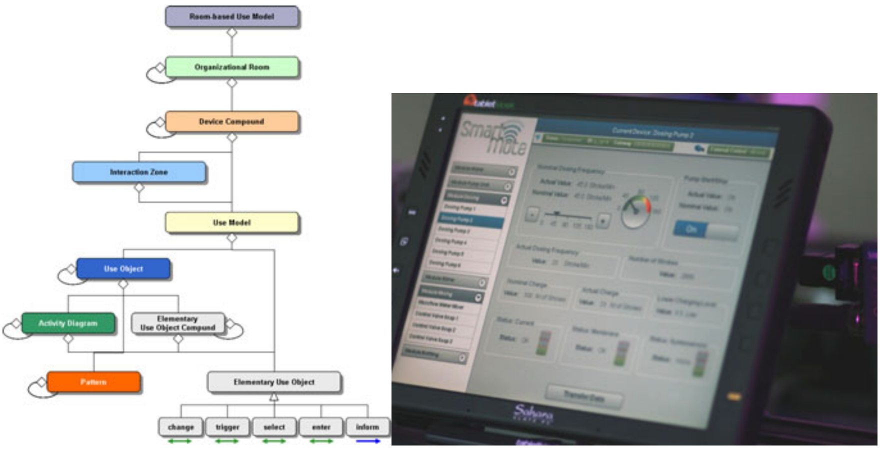


Fig. 1. Left: the room-based use model and its components. Right: the SmartMote running on a TabletPC used to access devices of the *SmartFactory*[KL].

2 SmartFactoryKL and SmartMote

Located in Kaiserslautern, the *SmartFactory*[KL] serves as a living lab for future ambient production environments [15]. It is characterized by a vast amount of field devices from various vendors. Basically, these devices build up a production process with the aim to produce colored soap filled up in dispenser bottles. According to the vendor, physical constraints, and device type these field devices offer diverse ways of human interaction – ranging from simple digit numbers to complex industrial workstations. Many of these devices possess their own user interfaces which – according to the vendor or physical device restrictions – provide different interaction concepts and can result in a bad user experience. The user has to adapt herself to different concepts, cooperate designs or modalities. Being confronted by this diversity during the execution of workflows this will result in operating errors [3].

Future production environments will consist of flexible entities which are replaceable and reconfigurable, but also aware of the entire process. In case of a device error during the production these lines will be able to reconfigure themselves in order to still be able to complete the production process. According to the current configuration the user has to adapt her mental model of the entire production flow in order to be able to locate the interface she has to access. For maintenance personnel this can be a challenging task.

Furthermore, the *SmartFactory*[KL] implements the paradigm of new ambient-intelligent production environments all of the field devices are interconnected and can provide their information at any time, which is a necessary requirement to develop universal interaction concepts. Given the combinatorial explosion of possibilities, traditional software development approaches are insufficient to develop such flexible user interfaces. Therefore, our universal interaction concept – the SmartMote – is based on various formal descriptions that can be altered and interpreted during runtime to produce a user interface giving access to the desired functionality of the field devices.

Core model of our approach is the room-based use model (RUM) which is built on the Useware Markup Language (useML) [5]. Figure 1 (left) depicts the RUM as it is used in our approach. It consists of a topological model of the entire environments comprising physical/organizational spaces (rooms) that can include devices. These devices are equipped with interaction zones describing the concrete restrictions on human-machine-interaction with respect to the current usage situation. Use model objects (UO) describe the tasks and workflows that can be executed by the user by a hierarchy down to elementary use objects (eUO) that stand for the individual user interaction such as change, trigger, select, enter and inform. Because the RUM only provides information about the content to be displayed there are many other aspects it does not cover. To interface to concrete application services the functional model is used [4]. We made use of the concept of usability patterns [2] to assign the elements to be displayed to concrete user interface elements in a presentation model.

In the following we will focus on the RUM – especially the contained task model – to show how the current usage situation has an influence on the content to be displayed and the effect on a user of the SmartMote.

3 Context in Ambient Intelligent Production Environments

Knowledge about the current usage situation can be used to improve the usability and perceived experience in various ways. During the interaction with the system the intention of the user can be predicted and the system can prepare itself to support future task executions. If the user will be within the operating range (e.g. during maintenance) of a robot arm while activating it this will lead to a hazardous situation. In case the system is able to detect the users' current position all functionality leading to this problem can be deactivated in advance and feedback can be provided to inform the user.

Context is a very broad term of which many definitions exist [12]. In general it summarizes all surrounding circumstances of a certain user interaction or activity. It contains every piece of information that can be used to characterize the environment in which the interaction takes place, attributes of the user as well of the devices the user is interacting with which additionally is relevant to this particular interaction [14]. Collecting this data in a formal model will create a knowledge base that can be used in order to derive facts about the current usage situation based on which the user interface can be adapted (e.g. deactivation of hazardous tasks). Thus, the information which is being provided to the user can be adapted to the current usage situation and the complexity of functionality offered to the user can be reduced to the functionality that is relevant in this situation and therefore usage errors can be prevented.

By the application of a set of rules on the contextual information will result in the generation of transformations as well as filters that will be applied to the RUM – certain tasks will not be available any more to the user. Therefore, the system needs to perform task reachability analysis in order to determine if the user can still reach her goal: completing the workflow. This aspect will be covered in section 5.

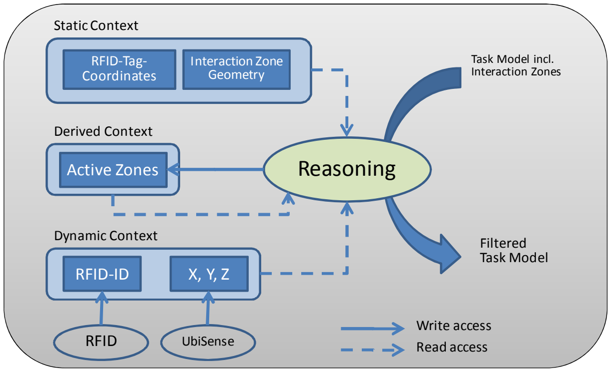


Fig. 2. Processing contextual information within the SmartMote – matching static and dynamic context factors to filter the RUM

In the following we will elaborate this idea based on the example of the SmartMote localized within the *SmartFactory*$^{\mathbf{KL}}$. As mentioned above the RUM comprises the specification of interaction zones, a concept initially developed for interaction in working environments [13]. According to the zone in which the user is currently located in, a different use model will be instantiated.

Figure 2 shows an overview of the context concept of the SmartMote. In the concrete scenario the indoor positioning system (UbiSense) as well as the RFID floor grid of the *SmartFactory*$^{\mathbf{KL}}$ will be used to determine the exact position of the user. Locations of the tags and geometric dimensions of the interaction zones are defined in the RUM. After receiving data from the positioning system and after contact to a RFID tag was established this information will be added to the RUM. This will result in firing a rule inside the reasoner component: given the information about the position of the user the SmartMote is able to match it to the dimensions of the single interaction zones as described in the RUM. In case a change of the active interaction zones was detected, a filter set can be generated in order to adjust the RUM, which will be interpreted and the displayed interface will be adapted accordingly to only show the functionality that was defined for the particular interaction zone. [1]

At this point the reasoner is only able to adjust the RUM in accordance to the contextual information derived from sensorial input. So far it was not possible to determine if the displayed user interface is still sufficient for the user to achieve her goals, which we will address with this contribution.

4 State of the Art

Before we elaborate our approach to check adapted user tasks, we want to give an overview about previous work on which our approach is built on. On the one hand this work involves task aspects as well as quality facets to check and modify task models. Therefore, in the following we will first briefly discuss task models before we introduce conceptual work on graceful degradation which is usually used in the domain of embedded systems.

4.1 User Task Models

Tasks that are required to be accomplished by the user in order to achieve her goal can be described in a hierarchical manner. Often this supports the stepwise refinement of the goal into abstract tasks down to atomic actions that will be executed during the interaction with the user interface. One of the earliest approaches to describe user tasks is the ConcurrTaskTrees (CTT) notation [8, 10]. Based on the ISO standard LOTOS (Language Of Temporal Ordering Specification) it leverages eight temporal operators to set the single tasks into relation with each other. CTT has constantly been improved but it shares one major drawback with other task models: scalability. Larger models can be created but confuse maintainers which can result in mistakes. Another disadvantage is the degree of refinement of basic tasks because the concrete type and nature of interaction is left open [9].

Based on the Useware Markup Language (useML), the RUM also structures tasks in a hierarchical manner [5]. An advantage of the RUM is the final refinement step towards leaf nodes in the graph structure – the elementary use objects (eUO). These describe the concrete atomic actions which are executed by the user. This makes the interpretation during runtime much easier.

4.2 Graceful Degradation

Graceful degradation is used in the domain of embedded systems to avoid failures in safety critical systems. The idea behind this concept is that a system consisting of diverse components is able to react to malfunctions of single devices and will compensate for the loss of e.g. actuators or sensors. Usually the system will be set into a different state in which it doesn't offer the full functionality anymore but it still is running the basic functionality. Graceful degradation describes the "smooth change of some distinct system feature to a lower state as response to errors" [11].

An example out of the automotive domain is the vitality of a car remaining functional even if subcomponents are not available anymore. In case of a failure of the brake booster the system will switch into a state were the car will only accelerate to a lower speed (than the vehicle's possible maximum). This concept can be transferred to the context of graphical user interfaces, where a failure within the 3D component could result in less graphical effects while still offering the basic functionality [11].

In model-based user interface development graceful degradation has been integrated in order to support the portability [6]. Since the user interface descriptions are available on different levels of abstraction, the portability between different platforms can be ensured much easier than in traditional development processes. Nevertheless, various platforms also provide their individual attributes and are e.g. restricted in terms of display size or resolution, which results in the natural question how to display the same content on a different resolution even if the modality remains the same. For example if the size of an interaction object is restricted graphical user interfaces can be implemented with different widgets that provide the same functionality (but with a loss of comfort, precision, etc.). In case the display space is limited the user interface can offer two buttons for selecting a value within a range (increase and decrease), but if the display space is more ample this function can be displayed as a slider. In the latter case the user can not only specify the value but she is also faster in giving her input by estimating the value on the scale before specifying the precise value. Thus, the concept of graceful degradation was used in this context by Florins [6] to address this problem in model-based user interface development by developing an algorithm, deriving a user interface according to certain platform constraints using a set of graceful degradation rules.

This inspired our work to apply the idea of graceful degradation directly to task models. Because dynamic aspects of ambient intelligent production environments will first have an effect on the tasks the user is allowed to perform it is a consequence to determine the new task sets and to exclude tasks which e.g. will lead into dead locks.

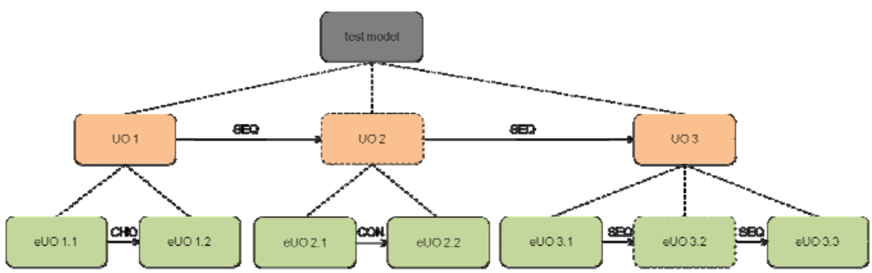


Fig. 3. A sample task model as contained by an RUM

5 Checking for Task Fulfillment by Graceful Degradation

Depending on the derived context, the resulting task model needs to be adapted accordingly. In the following we will introduce our approach based on a dialog model analysis in order to check for task fulfillment. First the underlying dialog model – which is generated out of the task model – will be introduced briefly. This serves as the input for the graceful degradation algorithm checking for task fulfillment and deriving an adapted task model.

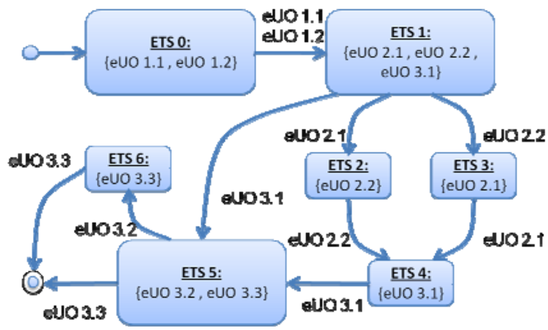


Fig. 4. Dialog model generated out of the task model shown in Figure 3

5.1 Task and Dialog Models

Based on a user task analysis, task models have been proven to be an effective way to formalize tasks that can be conducted by a user on a particular user interface in a hierarchical refined manner. User interfaces consist of dialogs which lead to the integration of a dialog model within our generation process to effectively generate user interfaces [7]. Starting with a task model as depicted in Figure 3, we automatically derive a dialog model in a two-phase process. Staring with a binary priority tree of the task model all possible execution orders need to be stored into a deterministic finite automaton. Enabled task sets (ETS) will be extracted from this binary priority tree by stepwise execution (after which the tree will be updated again) and determining the next ETS. Result will be an automaton (state machine) as required.

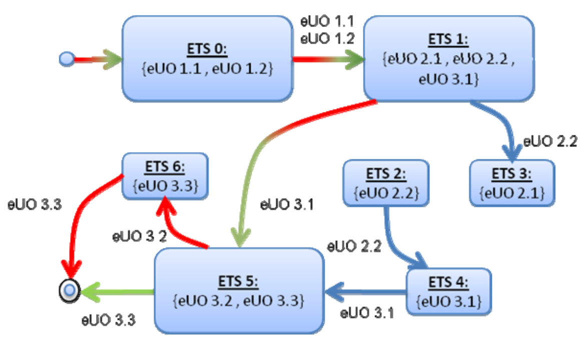


Fig. 5. Dialog model after eUO 2.1 becomes unavailable, removed transitions and marked transitions leading to the accept state

5.2 Task Fulfillment

After the dialog model has been generated it is possible to check if a user can still achieve her goal while using the adapted user interface (generated based on the new model). Figure 4 shows the resulting dialog graph of the task model shown in Figure 3.

Now, in case a service – and the corresponding elementary use object – becomes unavailable, all execution sequences within the model that are in any relation to the unavailable objects will be deleted.

In the given example the elementary use object 2.1 becomes unavailable. Therefore all transitions that are associated to the object will be removed in the automaton as depicted in Figure 5.

Checking all possible paths within this graph than can lead to the accept state the algorithm can determine if this state is reachable. As shown in figure 5, the algorithm will find two possible paths within this model that fulfill this requirement (red and green paths) and as a result declare the users goal as achievable. Four valid execution orders can be found in the example:

- eUO 1.1, eUO 3.1, eUO 3.3
- eUO 1.1, eUO 3.1, eUO 3.2, eUO 3.3
- eUO 1.2, eUO 3.1, eUO 3.3
- eUO 1.2, eUO 3.1, eUO 3.2, eUO 3.3

This is a necessary precondition in order to continue processing with the adaptation because if this check would be omitted the resulting user interface could be – with respect to its functionality – not useable anymore and therefore directly lead to operating errors.

All the elements of the initial task model will be set into the state *not reachable* to be processed by the graceful degradation approach. Starting with the leaf nodes – always elementary use objects – all these elements which are included in the valid execution orders determined from the dialog model analysis will be set to the state *reachable*. In case an object only contains children that are not reachable it will remain in its state. This algorithm will proceed recursively until the root node is reached (see Figure 6).

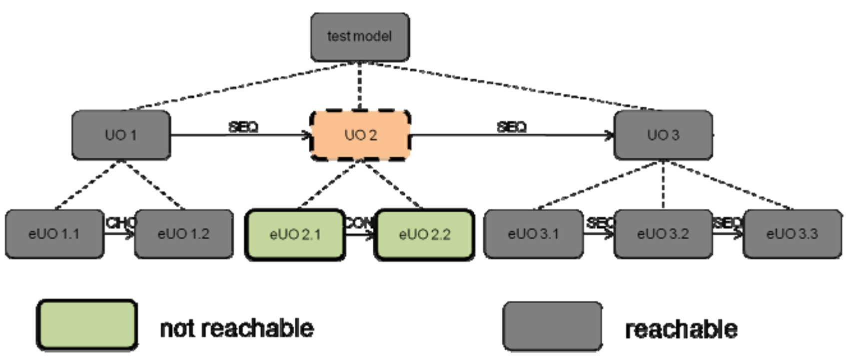


Fig. 6. Task model – marked according to reachability of tasks

Beginning with the root node, all the objects that are marked as not reachable will be removed from the task tree. After this phase is finished the adapted dialog model can be generated which serves as the input for the generated user interface.

5.3 Threats to Validity

First simulations have shown the feasibility of this approach, but also indicated some drawbacks.

Simulated based on models of the *SmartFactory*^{KL}, this approach works quite well (given the fact that it represents a small to medium size production environment), but using larger models will result in a lowered performance and if those models exceed a certain complexity the user interface will not be useable during run-time anymore.

Furthermore, processes within the *SmartFactory*^{KL} are modeled in a redundant way. This means in case of a malfunction of a pump there are other pumps to compensate its function. Therefore, these redundancies are also reflected within the task model according to the device configurations. Relating to the approach, the user has different ways to reach her goal and in case of an adaptation this can still be possible. If there are no redundancies, the user interface cannot be adapted but also the production environment will not be functional any more in our scenario.

6 Summary and Future Work

Ambient production environments will offer single holistic interaction concepts which offer interaction with all the deployed field devices and therefore will be able to provide a homogeneous user experience. One sample universal control device is the SmartMote which serves as a demonstrator and application scenario for our research.

In this paper we presented one step within our model-based user interface generation approach that is capable of adapting the user interfaces to contextual circumstances. Relying in task models describing the users' interaction with the interface, previous approaches were not able to determine the quality in terms of reachability of the users' goals of the resulting adapted user interface. We elaborated our approach based on this specific tool chain to leverage the determination of valid execution orders after a contextual chance. Based on the detected execution orders the task model can be adapted accordingly, avoiding tasks that cannot be executed by the user any more. Thus, also the resulting user interface does not offer any user interaction or dialogs which cannot be reached by the user or will lead to dead locks within the users' task execution process.

As mentioned above we successful simulated this approach with various models and adaptations based on our living lab while also indicating drawbacks. We therefore need to conduct further investigations on the performance of this algorithm and we are simultaneously developing additional algorithms against we will test the presented approach.

References

1. Bizik, K.: Context Modeling in the Domain of Ambient Intelligent Production Environments, Diploma Thesis, University of Kaiserslautern, Germany (2010)
2. Breiner, K., et al.: SmartMote: Increasing the Flexibility of Run-Time Generated User Interfaces by Using UIML. In: Proceedings of the 2nd International Workshop on Visual Formalisms for Patterns (VFfP 2010), located at the IEEE Symposium on Visual Languages and Human-Centric Computing 2010, Madrid, Spain (2010)

3. Breiner, K., Gauckler, V., Seissler, M., Meixner, G.: Evaluation of user interface adaptation strategies in the process of model-driven user interface development. In: Proceedings of the 5th International Workshop on Model-Driven Development of Advanced User Interfaces (MDDAUI-2010), located at CHI 2010, CEUR-Proceedings, Atlanta, GA, United States, April 10 (2010)
4. Breiner, K., Görlich, D., Maschino, O., Meixner G.: Towards automatically interfacing application services integrated in a automated model-based user interface generation process. In: Proceedings of the Workshop on Model-Driven Development of Advanced User Interfaces (MDDAUI 2009), 14th International Conference on Intelligent User Interfaces (IUI 2009), Sanibel Island, Florida (2009)
5. Breiner, K., Görlich, D., Maschino, O., Meixner, G., Zuehlke, D.: Run-time adaptation of an universal user interface. In: Proceedings of the 13th International Conference on Human-Computer Interaction (HCII 2009), San Diego, CA (2009)
6. Florins, M.: Graceful Degradation: A Method for Designing Multiplatform Graphical User Interfaces. PhD Thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium (2006)
7. Meixner, G.: Model-based Useware Engineering. In: Proc. of the W3C Workshop on Future Standards for Model-Based User Interfaces (MBUI 2010), Rome, Italy (2010)
8. Paternò, F.: ConcurTaskTrees: An Engineered Notation for Task Models. In: Diaper, D., Stanton, N. (eds.) The Handbook of Task Analysis for Human-Computer Interaction. Lawrence Erlbaum Associates, Mahwah (2003)
9. Paternò, F.: Tools for Task Modelling: Where we are, Where we are headed. In: Pribeanu, C., Vanderdonckt, J. (eds.) Proc. of the First International Workshop on Task Models and Diagrams for User Interface Design (TAMODIA 2002), pp. 10–17. INFOREC Publishing House, Bucharest (2002)
10. Paternò, F., Mancini, C., Meniconi, S.: ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models. In: Proc. of the 6th IFIP International Conference on Human-Computer Interaction (1997)
11. Saridakis, T.: Design Pattern for Graceful Degradation. In: Noble, J., Johnson, R. (eds.) Transactions on Pattern Languages of Programming I. LNCS, vol. 5770, pp. 67–93. Springer, Heidelberg (2009)
12. Strang, T., Linhoff-Popien, C.: A Context Modeling Survey. Workshop on Advanced Context Modeling, Reasoning and Management. In: 6th International Conference on Ubiquitous Computing (UbiComp 2004), Nottingham, England (2004)
13. Streitz, N.A., Röcker, C., Prante, T., Stenzel, R., van Alphen, D.: Situated Interaction with Ambient Information: Facilitating Awareness and Communication in Ubiquitous Work Environments. In: 10th International Conference on Human-Computer-Interaction (HCII 2003), Crete, Greece (2003)
14. Wurdel, M., Forbrig, P., Radhakrishnan, T., Sinnig, D.: Patterns for Task- and Dialog-Modeling. In: 12th International Conference on Human-Computer Interaction (HCII 2007), Beijing, P.R. China (2007)
15. Zühlke, D.: SmartFactory - A Vision becomes Reality. Keynote Paper of the 13th IFAC Symposium on Information Control Problems in Manufacturing, INCOM 2009 (2009)

Towards an Automatic Analysis of Interaction Data for HCI Evaluation

Application to a Transport Network Supervision System

Selem Charfi[1], Houcine Ezzedine[1], Christophe Kolski[1], and Faouzi Moussa[2]

[1] Univ Lille Nord de France, F-59000 Lille, UVHC, LAMIH, F-59313 Valenciennes, CNRS, Fre 3304, F-59313 Valenciennes, France
[2] Laboratoire Cristal, Ecole Nationale des Sciences de l'Informatique de Tunis, rue des Entrepreneurs Charguia II, 2035 Tunis-Carthage, Tunisia
CharfiSelem@gmail.com,
{Houcine.Ezzedine,Christophe.Kolski}univ-valenciennes.fr,
FaouziMoussa@gmail.com

Abstract. In this paper, we propose a method for the analysis and the interpretation of the interactions between the user and the interactive system to evaluate. The proposed approach is based on the comparison between the user interaction sequences and the sequences average. This confrontation concerns the task execution duration, the realized tasks number and the action sequence used in a defined experimental scenario.

Keywords: Human-Machine Interaction, Automatic Evaluation, Ergonomic Guidelines, Task modeling.

1 Introduction

For more than thirty years, many researches concerned the improvement and the optimization of different interaction aspects between human and machines. This interest increases considerably with the interactive systems evolution. The HCI evaluation can be done in every development phases (specification, design, implementation, tests…) [22]. In this paper, we are interested in the evaluation. To evaluate the interactive systems, many methods, techniques and tools exist. Nevertheless, they often have their own requirements and characteristics, and fill into a category as ergonomic inspection, verbalization, electronic informer, questionnaire, etc. The application of the same method, technique or tool, can give various results with various users and/or various involved evaluators [5], [16]. In the case of the approaches that allow the interaction analysis, it is sometimes very difficult to study their result. Indeed, usually the evaluator faces an important amount of data [4]. Its treatment generally requires a very long time and a great effort to interpret it. Thus, we propose an automatic interaction data analysis method.

In this article, we first present briefly the HCI evaluation. Secondly, we propose our approach for the HCI evaluation based on the analysis of the user interactions. Then, this proposition is applied on a transport network supervision system.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 175–184, 2011.
© Springer-Verlag Berlin Heidelberg 2011

2 Related Work

Generally the HCI evaluation is considered as measuring interactive system usability and the utility [9], [13], [16] [19] [22] [23].

A system is said useful if it allows the user to reach his or her work's objective [1] [15] [16] [20]. The usability concerns the learning and the usage ease of the system [10]. According to [22], the evaluation is a comparison of an object model to a referential model to establish conclusions. The object model is the examined system to be evaluated. This model can be defined by a set of interface attributes, quantitative and/or qualitative measures, observations… The referential model is specified by the evaluator and/or the experts. It constitutes an ergonomic reference. The comparison between these two models indicates possible defects and/or suggestions in order to improve the evaluated interface.

Several tools are based on this principle in the evaluation process. For instance, we can quote MM WebRemUSINE [17], MESIA [25] and EISEval [27]. These tools define the referential model as a selected action sequence by the evaluator and the object model as the interactions captured between the user and the system. This referential model is difficult to be established. Indeed, it requires advanced knowledge from the evaluator in several domains: software ergonomics, cognitive psychology, etc., as well as usability standards. Furthermore, to establish an adequate model, the evaluator has to have a good understanding of all the user profiles and the practicable tasks by the system.

3 Automatic Interaction Data Analyses

Our method consists in the capture, the analysis and the interpretation of the interaction data between the user and the interactive system. First, the method consists in capturing the interactions between the users and the system to be evaluated. Then, a Petri Net (PN) is generated. It models the captured interactions (PNs are often used in HCI, see for instance [6], [18]). This PN allows deducing the users' average sequence. Finally, the evaluation is done according to the differences founded between the realized action sequences and the average of these sequences, Fig. 1.

Our evaluation approach is structured as following. First, an object model is defined. It contains executed actions by the users during an experimentation scenario. We introduce the evaluation as the study of the difference between the various object models and the average of these models, Fig. 1.

The executed actions and tasks can be detected by several interaction capture devices: electronic informer [27], eye-tracking [2], observation, video [18]. A scenario is elaborated. It consists on executing, during the experimentation, the tasks by a set of users. Our approach consists in 3 steps, Fig 1. First, the interactions between the user and the interface have to be structured to a common format. Then, an average model resulting from all the object models is generated. Finally, the various object models are compared with the average model to evaluate ergonomic quality of the interface.

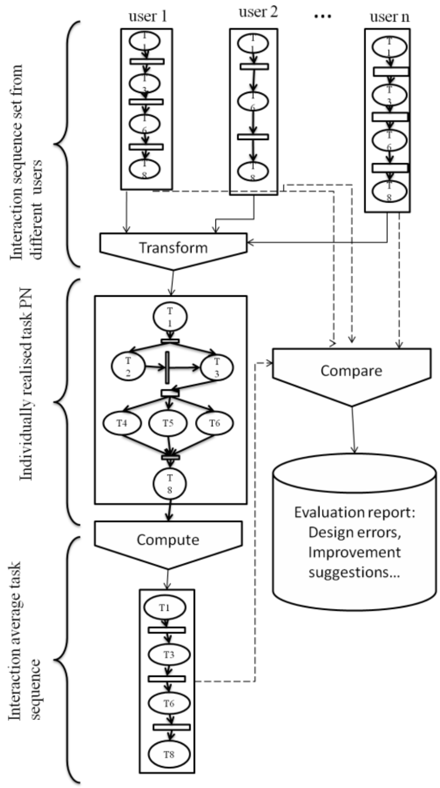


Fig. 1. Proposed evaluation approach

3.1 Interaction Modeling Using the Petri Nets

Several interaction modeling exist [11]. In our proposition, we adopt the one proposed by Tran [26]. This modeling consists in distinguishing the various events concerning the interfaces in two categories:

- - EVIU (*Evènement Interface Utilisateur* (in French), translated by: UI event): it consists on the events concerning the interface components: opening a window, exit a window, error message, form validation …
- - EVDI (*Evènement Dispositif d'Interaction* (in French), translated by: DI event): This set is related to the events of lower level which constitute the EVIUs. They handle directly the events generated by the interaction devices: right click on the mouse, the keyboard touch...

During the interaction, the user proceeds to several tasks. A task consists of one or several EVIU(s). An EVIU consists also of one or several EVDI(s), Fig. 2. The EVDI is codified in an XML file. Each EVIU is composed of an EVDI set. From this composition, we can easily detect the realized action sequences by the users during the experimentation process. These interaction sequences are illustrated as a PN.

These Petri Nets are coded in XML files to facilitate their storage and their management. Each place is coded using the name, previous place(s) and the following place(s). The transitions are coded using the passage's condition (and, or, exclusive or, etc.).

3.2 Average Object Model Calculation

Once the interaction file is generated (Fig. 3), a PN is elaborated. It includes the various interaction sequences as well as the probability to execute each task. This probability is calculated according to the percentage of the users who execute it. Places in this network are the tasks realized by the user. The transitions between these places give the numerical values of this probability, Fig. 3.

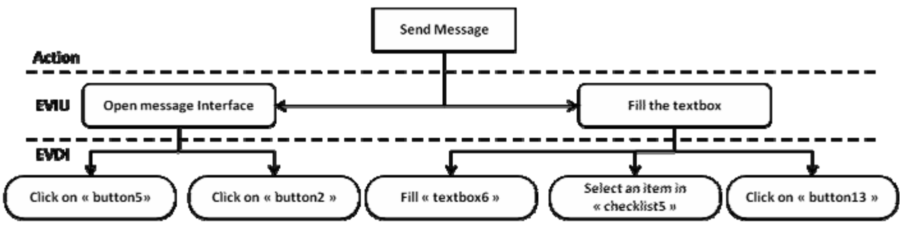


Fig. 2. Example of task decomposition into EVIU and EVDI

A reduction algorithm is applied to the resulting PN (repetition deletion, transitions buckles, etc.) to facilitate its interpretation. Once each PN is generated, the average PN is calculated. It is calculated by the enumeration of the various possible paths on the PNs and their respective probability. The average sequence is the one having the highest probability indicator. The sequence execution duration is the average of the sequences execution time. The task number average is calculated in the same way.

3.3 Study of the Difference between the Average Object Model and the Different Object Models

The difference between the object models and the average object model can be established on 3 levels: (1) actions number, (2) run time, (3) and gap between different action sequences performed by the user. If the gap is not important, we consider the users proceeded according to the same manner and the information is well presented. In that case, the users opted for the same action sequences in the experimentation scenario. In the other case, some ergonomic inconsistencies can be detected through the difference type (in the sense of the ergonomic criteria proposed by Bastien and Scapin [3] or proposed by the ISO 9241 standard). Problems met by the user during the experimentation can be detected. In the case of a runtime difference, it is possible to conclude that the user found difficulties while using the interface. If some users manipulated the system slowly, this can be due to problems in relation to the following criteria: *significance*; *workload*. In case of an important difference in the transitions number in the action sequences use, it becomes possible to show that the users proceeded to useless manipulations. This is related to the ergonomic criteria: *homogeneity/coherence*; *prevention against the errors*. When there's an important difference in actions sequences, the evaluated interface should be revised according essentially to the following criteria: *guidance* and *prevention against the errors*.

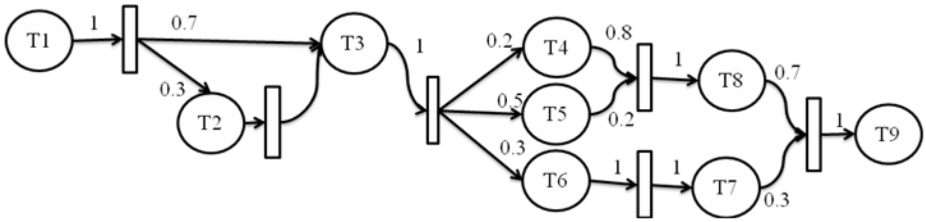


Fig. 3. Example of PN with the probability of each transition

We have performed a first experimentation involving 100 users. These users interact with a ticket machine system. The interaction sequences are captured from an electronic informer. The results are generated from a database, in which the electronic informer stores the different actions. The results show that there is a difference of +- 17% in the execution duration and a difference of +- 14% in the number of action.

4 Application to a Transport Network Supervision System

Our method is applied on an interactive system. This system is similar to the one used for the network transport supervision in Valenciennes (France). It is called the IAS: Information Assistance Support for Passenger. The IAS is an agent based system allowing to visualize the transport network [7] [8] [25], Fig. 4. The proposed approach was experimented in laboratory on 12 users. A phase of training was proposed to these users on the IAS (application case, task, priority, explanation, etc.). These users are not experimented in this task (the network supervision). They have never manipulated such systems. We chose novice users in order to better detect utility and usability problems related to the IAS.

The evaluation scenario consists of a particular situation. Indeed, the transport network is subjected to disturbances. The supervision task is not done in a collaborative way. During the experimentation, the users are submitted to particular circumstances which oblige them to inspect the whole tasks of the system. The interactions between these users and the IAS were captured by an Electronic Informer, called *EISEval* (Environment for Interactive System Evaluation) [27]. The results indicate that the system needs to be approved. In this evaluation, the vehicles passage schedules is disturbed. The user has to act via the IAS by sending messages (1) to the driver in the aim of accelerating either slowing down the vehicle and (2) to the passengers to warn them of a possible timetable change. In the selected scenario, there is an overlapping of several immediate delays/advances on one hand and the appearance and the disappearance of vehicules delay / advance on the other hand. Once the user interactions are captured, we establish the links between the events of user interface (EVIU) and the tasks realized by the system. These links inform the evaluator about the tasks sequence made. The sequence of interaction is generated for each user. The users proceeded on average to 39 elementary tasks. The evaluation scenario lasted on average 17 mn 32 s. From the PN (Fig. 5.a), the average PN can be deducted (Fig. 5.b) by calculating the action sequence the most probable. The comparison between the action sequences and the average PN (Fig. 5.b) shows that there is an important difference between user actions.

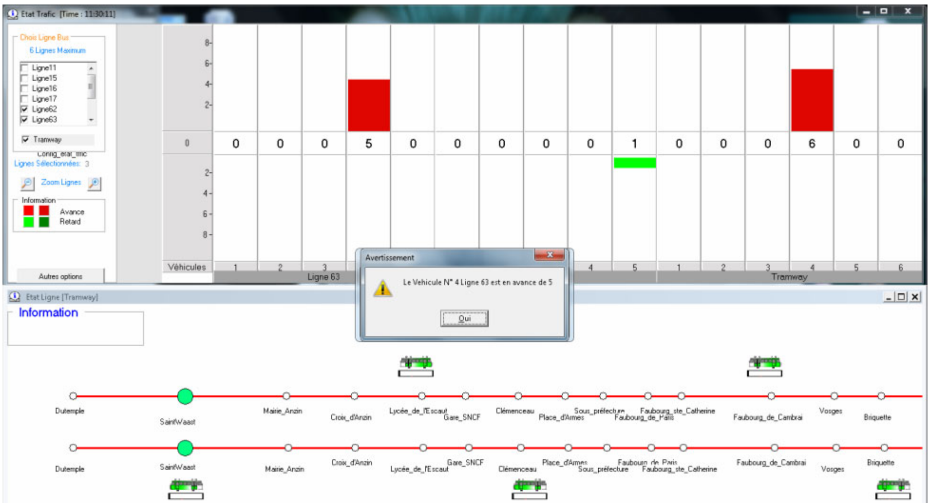


Fig. 4. Extract of the evaluated interface (Part of the IAS)

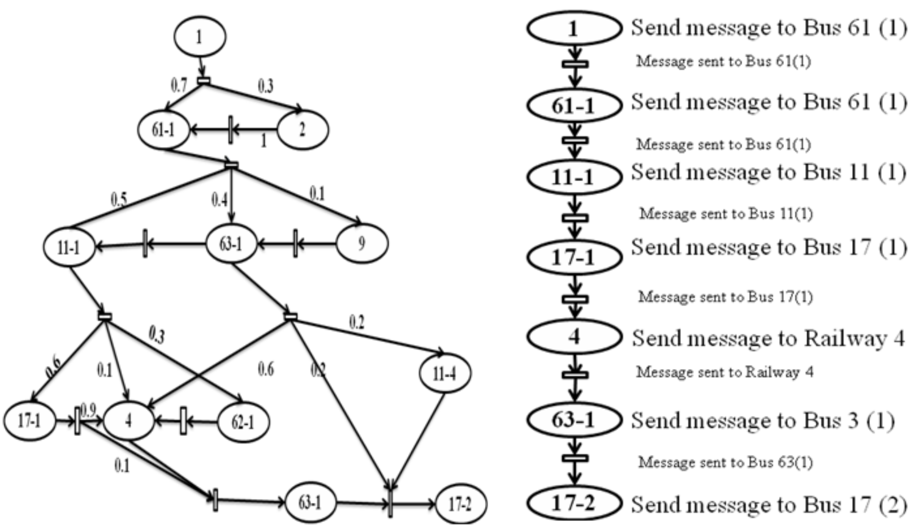


Fig. 5. Example of PN deduced in the experimentation

This difference is related essentially to generated events. Indeed, the users 3, 7 and 8 proceeded useless manipulations (more than 30 %), Table 1. From the action sequences, we notice that numerous actions were repeated usefully. This proves that there are problems according with the *coherence*, *guidance* and *prevention against errors* criteria. In these situations, the users were not sure to have executed an action. To improve the interface, we propose to add a new agent to the IAS; this agent will indicate the user executed action history to avoid repetitive actions.

We find some problems related to the prevention against the errors criteria. The calculated difference in the task run time is not very important (more or less one minute). The IAS presented information is very clear to the user. The ergonomic guidelines related to the" information display" and the "data posting" are considered as respected in the interface.

Table 1. Extracts of results issued of the experimentation

User	Execution Time	Time space	Realized actions	Task space
2	14:07	-19.48	36	-07.69
3	17:21	-01.04	27	-30.76
4	17:04	+03.04	48	+23.07
5	19:24	+01.64	42	+07.69
7	18:49	+07.31	22	+43.58
8	17:24	-00.76	42	+07.60

Thanks to the experimental evaluation, certain criticisms and suggestions are emitted to improve the IAS, Fig. 5. First of all, to warn against possible errors, the user has to confirm every time his or her actions via a dialog box containing the message that he or she wishes to send, the address, a validation button and a cancel button. On the other hand, some users proceeded to repetitive actions. This is due to the fact that they were not sure of executing some tasks. Furthermore, some certain

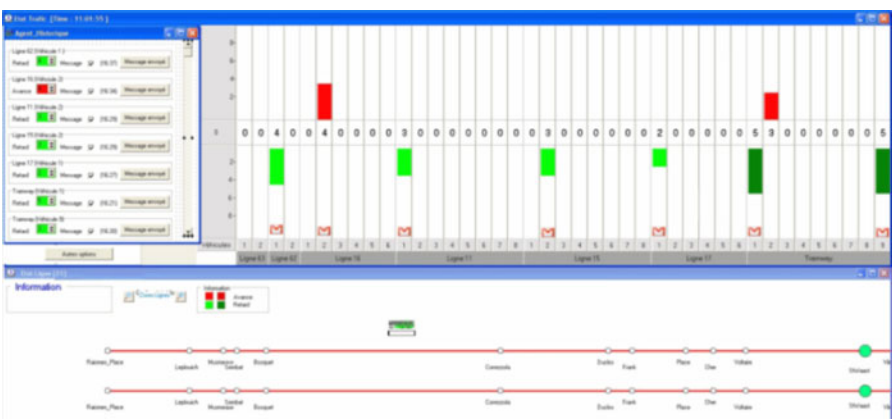


Fig. 6. The interface suggested to be added to the IAS (extract)

forgot if they handled certain disturbances or not. A new agent is suggested to be added to the IAS. This agent task is to illustrate the executed action history. Furthermore, we suggest adding on the agent “Traffic state”, an icon as small indicator every time the user sends a message to a vehicle, Fig 6.

5 Discussion

In this evaluation, several difficulties concerning the IAS use were discovered. Although the users belong to the same profile, we found important differences in the number of actions during the experiment. This is essentially due to the non compliance with the *prevention against the errors* and *guidance* criteria. The found results are in agreement with those obtained by Tran [26] and Trabelsi [24] for the IAS evaluation. Tran’s results show that the users often proceed to useless manipulations. In our approach, we did not specify a referential model. It was specified from the user interaction sequences with the evaluated system.

The proposed approach has been applied to the evaluation of an information assistance system. Several improvements of the interface are suggested. This is mainly in the addition of a new agent at the interface. We consider than our approach is easier to use than EISEval [27]. PN, generated from the interaction, are easier to perform, interpreter and exploit. Nevertheless, the proposed approach allows only assessing interfaces according with classical ergonomic criteria. It does not consider new criteria. In addition, the approach is automatic only in capture and analysis phases. The criticism and the suggestions are established manually (according to the classification of evaluation tools by Ivory and Hearst [12]). Additional methods or tools should be proposed to improve the ergonomic diagnosis delivered by the system; it is possible for example to use either questionnaires or the Cognitive Walkthrough method [13]. Furthermore, it would be interesting to integrate a direct observation of the users during the tests. This would facilitate the understanding and the explanation of certain ergonomic inconstancies.

6 Conclusion

An approach for the study and the analysis of the human-machine interactions was presented in this paper. This approach allows supporting interactive system evaluation. It aims present criticisms and suggestions to the evaluator to improve the evaluated interface.

The proposition has been applied to a network transport supervision system. We propose to apply it on other interactive systems. In addition, we suggest to extend it by combining this approach with other evaluation methods (for instance questionnaire and usability inspection). This would allow validating the ergonomic criteria to consider during the evaluation. We plan also to integrate the resulting approach into a generic, flexible and configurable environment for the evaluation of interactive systems.

Acknowledgement. The present research work is partially supported by the International Campus on Safety and Intermodality in Transportation (CISIT), the Nord-Pas-de-Calais Region, the European Community, the Regional Delegation for Research and Technology, the Ministry of Higher Education and Research, and the CNRS.

References

1. Abran, A., Khelifi, A., Suryn, W., Seffah, A.: Usability Meanings and Interpretations in ISO Standards. Software Quality Journal 11, 325–338 (2003)
2. Baccino, T.: Oculométrie Cognitive. In: Tiberghien, G. (ed.) Dictionnaire des Sciences Cognitives, pp. 100–101. Armand Colin, Paris (2002)
3. Bastien, J., Scapin, B.: Ergonomic Criteria for the Assessment of Human-Computer interfaces. INRIA report no 156, Le Chesnay (1993)
4. Bastien, J.M.C., Scapin, D.L.: Les méthodes ergonomiques: de l'analyse à la conception et à l'évaluation. In: Drouin, A., Robert, J. (eds.) Ergonomie et Informatique Avancée, Ergo-IA 2002. I.D.L.S, Biarritz (2002)
5. Charfi, S., Dubois, E., Scapin, D.: Usability Recommendation in the design of mixed interactive systems. In: Proceedings of the 1st ACM SIGCHI symposium on Engineering Interactive Computing Systems, Pittsburgh, PA, USA, pp. 231–236 (2009)
6. Ezzedine, H., Abed, M.: Une méthode d'évaluation d'Interface Homme-Machine de supervision d'un procédé industriel. Journal Européen des Systèmes Automatisés (JESA) 7, 1078–1110 (1997)
7. Ezzedine, H., Kolski, C., Péninou, A.: Agent oriented design of human- computer interface. Application to Supervision of an Urban Transport Network. Engineering Applications of Artificial Intelligence 18(3), 255–270 (2005)
8. Ezzedine, H., Trabelsi, A., Kolski, C.: Modelling of an interactive system with an agent-based architecture using Petri nets, application of the method to the supervision of a transport system. Mathematics and Computers in Simulation 70(5-6), 358–376 (2006)
9. Farenc, C.: ERGOVAL: une méthode de structuration des règles ergonomiques permettant l'évaluation automatique d'interfaces graphiques (in French). PhD Thesis, University of Toulouse 1, France (1997)
10. ISO. ISO/IEC 9126-1: Software engineering, Product quality, Part 1, 2, 3 and 4, (JIS X 0129)
11. Hilbert, D.M., Redmiles, D.F.: Extracting usability information from user interface events. ACM Computing Surveys 32(4), 384–421 (2000)
12. Ivory, M., Hearst, M.: The State of the Art in Automated Usability Evaluation of User Interfaces. ACM Computing Surveys 33(4), 173–197 (2001)
13. Lazar, J.: Universal Usability: Designing Computer Interfaces for Diverse User Populations. John Wiley and Sons, Chichester (2009)
14. Mahatody, T., Sagar, M., Kolski, C.: State of the Art on the Cognitive Walkthrough method, its variants and evolutions. International Journal of Human-Computer Interaction 26(8), 741–785 (2010)
15. Navarre, D., Palanque, P., Winckler, M.: Model-Based Evaluation: A New Way to Support Usability Evaluation of Multimodal Interactive Applications. In: Law, E., Thora Hvannberg, E., Cockton, G., Vanderdonckt, J. (eds.) Maturing Usability: Quality in Software, Interaction and Quality (chapter 5). Human-Computer Interaction Series, pp. 96–122. Springer, Heidelberg (2007)

16. Nielsen, J.: Usability Engineering. Academic Press, Boston (1993)
17. Palanque, P., Bastide, R.: Interactive Cooperative Objects: an Object-Oriented Formalism Based on Petri Nets for User Interface Design. In: Proc. of the IEEE SMC 1993, Systems Engineering in the Service of Humans, Le Touquet (October 1993)
18. Paterno, F., Piruzza, A., Santoro, C.: Remote Web usability evaluation exploiting multimodal information on user behavior. In: Proc. CADUI, Bucharest, Springer, Heidelberg (2006)
19. Rubin, J., Chisnell, D., Spool, J.: Handbook of usability testing: How to plan, design, and conduct effective tests, 2nd edn. John Wiley and Sons, Chichester (2008)
20. Seffah, A., Donyaee, M., Kilne, R.B., Padda, H.K.: Usability measurement and metrics: A consolidated model. Software Quality Journal 14(2), 159–178 (2006)
21. Senach, B.: Évaluation ergonomique des IHM: Une revue de la littérature (in french). INRIA report no 1180 (March 1990)
22. Scapin, D.L., Bastien, C.: Analyse des tâches et aide ergonomique à la conception: l'approche MAD*. In: Kolski, C. (ed.) Analyse et Conception de l'IHM. Interaction Homme-Machine pour les SI, Hermès, Paris, vol. 1, pp. 23–49 (2001) ISBN 2-7462-0239-5
23. Stone, D., Jarrett, C., Woodroffe, M.: User interface design and evaluation. Morgan Kaufmann/Elsevier, San Francisco (2005)
24. Trabelsi, A.: Contribution à l'évaluation des systèmes interactifs orientés agents: application à un poste de supervision du transport urbain (in french). PhD Thesis, University of Valenciennes and Hainaut-Cambrésis, Valenciennes, France (2006)
25. Trabelsi, A., Ezzedine, H., Kolski, C.: Evaluation of agent-based interactive systems, application to an information assistance system: first results. In: Sayed-Mouchaweh, M. (ed.) 28th European Annual Conference on Human Decision-Making and Manual Control, Reims, France, pp. 45–50 (September 2009)
26. Tran, C.: Vers un environnement générique et configurable pour l'aide à l'évaluation des systèmes interactifs à base d'agents (in french). PhD Thesis, University of Valenciennes and Hainaut-Cambrésis, Valenciennes, France (2009)
27. Tran, C., Ezzedine, H., Kolski, C.: Evaluation of Agent-based Interactive Systems: Proposal of an Electronic Informer Using Petri Nets. Journal of Universal Computer Science 14(19), 3202–3216 (2008)

A Formal Model of Mixed-Initiative Interaction in Design Exploration

Sambit Datta[1] and Michael Hobbs[2]

[1] School of Built Environment,
Curtin University
sambit.datta@curtin.edu.au
[2]School of Information Technology,
Deakin University
mick@deakin.edu.au

Abstract. Computer-based environments for supporting design are complex software artifacts. These tools need to use sound computational formalisms as well as address issues of human usability. The development of interactive and usable generative systems is a significant research area in design computation. Though classical search techniques play a central role in the generative kernels of these "closed-world" systems, the open-ended exploration of design spaces is the desirable goal. In this paper, we present a formal model of exploration that combines search with user driven exploration. We describe the role of interaction and agency in an experimental mixed-initiative design support system.

Keywords: generative design, mixed-initiative, design exploration.

1 Background

Generative design comprises an iterative process of specifying problems, finding plausible and alternative solutions, judging the validity of solutions relative to problems and reformulating problems and solutions. Mixed-initiative is an effective paradigm for addressing the process of directing problem-solving goals (Cohen et al., 1998) in a domain of discourse. Through mixed-initiative, the user and the formalism can share responsibility over domain goals. For example, Rich and Sidner (1997) and Rich and Sidner (1998) demonstrate a domain level collaboration through an interface agent that works on a plan with its user. Veloso (1996) and Veloso et al. (1997) employ a shared representation in the planning domain. Both automated and human planners are able to interact and construct plans jointly. Smith and Hipp (1994) propose a common meaning representation to achieve goals in natural language dialogue through mixed-initiative. Guinn (1996) considers initiative over mutually shared goals and how goals are solved by the participants (agent and human) in spoken dialogue systems.

Mixed-initiative over a domain goal requires both humans and automated software to share a representation of the domain of discourse. Designers share tasks with the formalism through an interaction model that connects the designer's view of the domain with the symbol level constructs available for computing exploration. From the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 185–193, 2011.
© Springer-Verlag Berlin Heidelberg 2011

designer's perspective, the representation of the domain must account for and connect onto the concepts underpinning the design space formalism. A difficulty of explanation arises in this task: the elements of the domain layer collapse into and find explanation in the sparse symbol-level machinery below. One formal device in the substrate serves several concepts in the domain layer. To address these difficulties, it is necessary to maintain three levels in the exposition of the domain layer: the designer's view of the components of exploration, the formal substrate underpinning these views and finally, domain layer concepts that map the user level concepts onto the formal components of the design space formalism.

2 Interaction with a Description Formalism

The description formalism described in Woodbury et al. (1999), provides a formal substrate for supporting the entities of exploration, state, move and structure. The symbol substrate for design space exploration is reported in Woodbury et al. (1999). It implements a formal mechanism for computing exploration in terms of types, features, descriptions and resolution algorithms (Burrow and Woodbury, 1999; Woodbury et al., 2000). The formalism is based on typed feature structures (Carpenter, 1992) and extensions (Burrow, 2003). In particular, the representational properties of information ordering, partiality, intensionality, structure sharing and satisfiability are addressed in the formalism. To define how designers may employ the entities of the substrate at the user level, a model of interaction, as shown in Figure 1 is necessary.

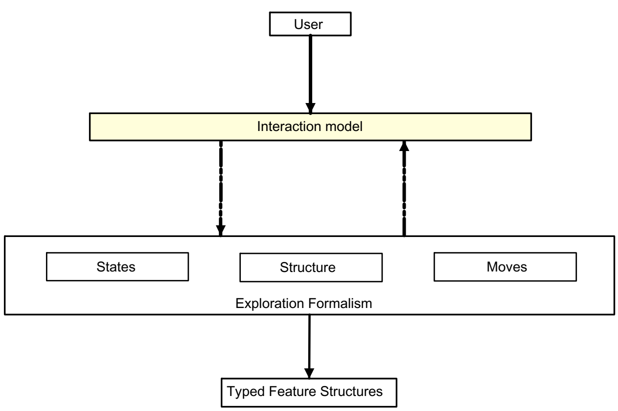


Fig. 1. An interaction model integrates the user and the description formalism

Mixed-initiative presents a paradigm, for combining a human designer and a description formalism through the specification of communication, coordination and control of the exploration process. To address the above, a three-layered model is developed for interactive exploration. Through the interaction model, both user and formalism must have the flexibility to acquire or relinquish initiative during exploration. The layers of the interaction model are shown in Figure 2. Each layer plays a distinct role for addressing the requirements for mixed-initiative exploration.

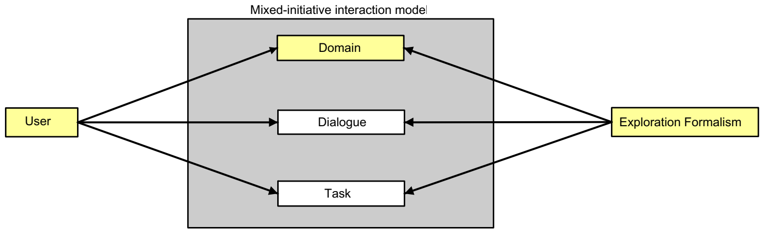


Fig. 2. A mixed-initiative interaction model integrates the user and the description formalism

The domain layer bridges the gap between the knowledge level formulation of the designer's view of the exploration domain and the symbol level substrate of the description formalism. The requirements of the domain layer in the mixed-initiative model of exploration are:

Support the designer's view of the domain. The domain layer must support the designer's view of the domain of exploration, namely, problems, solutions, choices and history. The domain layer provides concepts for the representation of problems, their reformulation and the generation of alternative solutions from the user's perspective. The domain layer mediates between the designer's view of exploration comprising problems, solutions, choices and history; and a design space representation aimed at efficient generation, indexing and recall.

Support joint responsibility over goals. Supporting joint responsibility over domain goals is a major problem in design exploration. Through mixed-initiative, the domain layer enables both the designer and the formalism to maintain context and share responsibility over goals in the domain of exploration.

3 Implementation

The designer's view of exploration comprises an account of problems, solutions, choices, their connections and the developing space of explicitly discovered design alternatives. This view is less concerned with the formal specification of internals and more with the existence of objects and the external hooks necessary to support interactive exploration. The designer's model of exploration comprises problems, solutions, choices and history (their connections and the resulting explicit design space). The problem formulation and reformulation cycle, the solution generation and reuse cycle, the intentional choices of the designer and the rationale of exploration in the form of a history are captured in this view. The representation of the designer's view is shown in Figure 3.

The structure of exploration is represented through the ordering relation of subsumption. The concept of an ordered design space underpins the description formalism. Exploration states are ordered by the relation of subsumption, a formal ordering over the collection of states. Burrow and Woodbury (2001) argue that design history is a significant device for supporting exploration. Choices, their connections

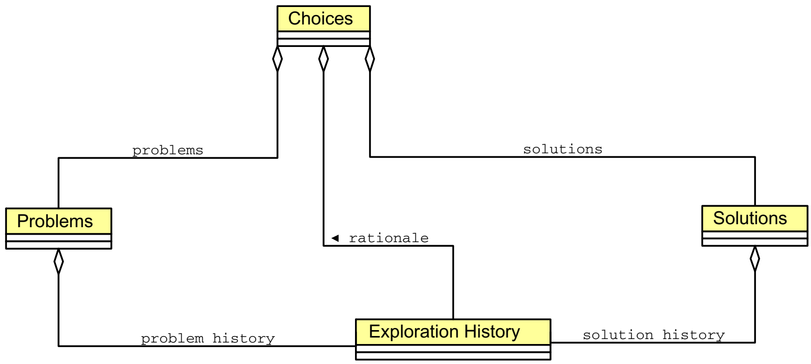


Fig. 3. The designer's view of exploration can be captured through a representation of the following entities: problems, solutions, choices and history. The problem formulation and reformulation cycle, the solution generation and reuse cycle, the intentional choices of the designer and the rationale of exploration in the form of a history are captured in this view.

and the developing history of explicitly discovered design alternatives must be accessible to the designer through interaction with the structure of exploration. The interaction model provides the designer with a view of design space structure. From the designer's perspective such a model must captures the elements of the history of design exploration.

The designer's view can be identified with the entities described in Section 2. It is necessary to explain these at a second level: that answers to both the designer's model and to the formal substrate. The mapping from the designer's view of exploration to the symbol structures of the underlying machinery is made explicit through four constructs in the domain layer, namely, problem state, solution state, choice and satisfier space. Figure 4 shows the mapping of the designer's view of exploration to the constructs in the domain layer.

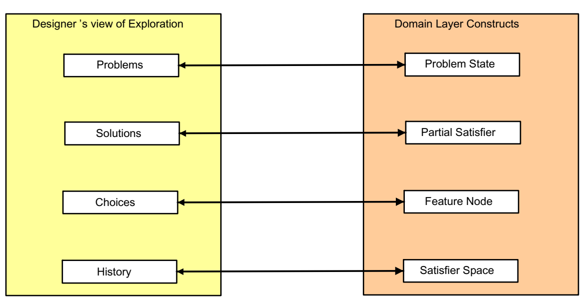


Fig. 4. Mapping the designer's view of exploration to constructs in the domain layer

Problems become problem states. Solutions become partial satisfiers. Choices compose a relation between problem states and partial satisfiers. Problems, solutions and choices recording the history of exploration are captured by satisfier space. Each of these constructs are developed in greater detail, following a three level exposition, the designer's view, the symbol substrate, and the mapping of the two based on the domain layer as shown in Figure 4.

3.1 Problem State, Pstate

To a designer, problems comprise both design requirements and desired properties of an artifact. Problems may be hierarchical, that is in addition to requirements and properties they may comprise sub-problems, which themselves may be similarly recursive. Designers revise problems as aspects of a design situation reveal themselves through exploration, the conception of the actual problem being solved may change. In a designer's problem space, work is done by a combination of problem formulation (specifying/adding) requirements, attributes and sub-problems) and problem revision (removing or modifying the same).

At the symbol substrate, problems are specified through two main constructs of the typed feature structure mechanism: a type hierarchy <**Type,** ≤ > and a description drawn from Desc. First, the specification of a problem amounts to the construction of an inheritance hierarchy of types. The types are refined by a collection of features introducing appropriateness constraints on types. Second, a problem to be solved is expressed using descriptions drawn from Desc with respect to <**Type,** ≤ >. Present here is the representation of problems through type hierarchy construction and the authoring of descriptions. Detailed expositions of how these structures are supported in given in Woodbury et al. (1999).

As shown in Figure 5, the domain layer construct, Pstate, maps to the type hierarchy and to descriptions. From the perspective of an exploration process, a design problem can be expressed as a type inheritance hierarchy. Typically these would be such forms that have been ossified by past experience in design space. In this case, the problem would be available as a problem state, specified by its trivial description as a type in the type hierarchy. The mapping of a problem state to the type system can be explained as follows:

$$\text{PState} \equiv \bullet\text{PState} : \text{InheritanceHierarchy} : \text{Type} \quad (1)$$

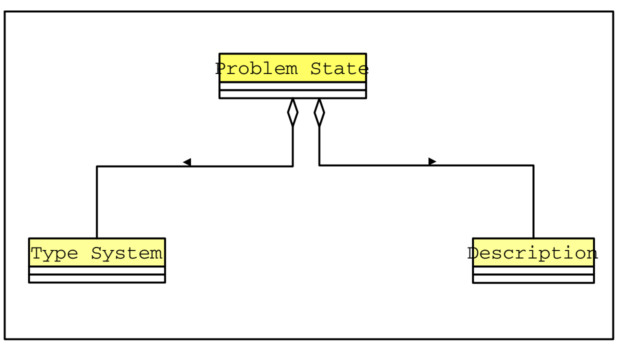


Fig. 5. The problem state composes a collection of descriptions, Desc over a type system

A problem can also be specified as a description specifying certain forms of spatial relations or constraints on types based on the description language. The connection between a PState and the formal substrate of descriptions drawn from Desc is given as follows:

$$\text{PState} \equiv \bullet\text{PState} : \text{Desc} : \text{Description} \tag{2}$$

In the above case, the exploration of the initial description would amount to designer interaction with the the domain layer construct, a PState. Through interaction with a PState, the designer can define design requirements either within the type system or through a collection of descriptions. Figure 6 extends Figure 5 by expanding a problem state to reveal its connections with typed feature structures. Summarising, the problem state is defined as a type or a description over a hierarchy of types. The domain layer construct Pstate, encapsulates the two distinct views of the problem formulation and reformulation process in the combined representation.

3.2 Solution States, SState

In the domain of design, problems may have no solutions, a finite number of solutions or an arbitrarily large collection of solutions. For example, in the SEED Knowledge Level, a solution is modelled as a "design_unit" (Flemming and Woodbury, 1995), representing a physical and geometric model. To a designer, a solution is a component in the spatial or physical structure of a building and has an identifiable spatial boundary. Thus solutions describe physical and geometric characteristics of structures satisfying problem descriptions.

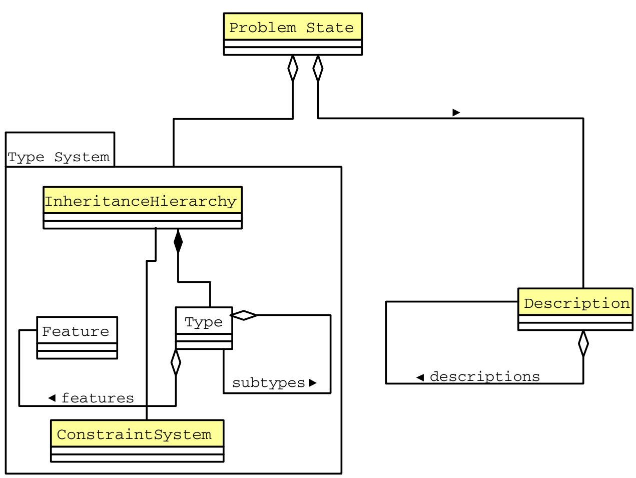


Fig. 6. Types, Features, constraints and descriptions comprise the representation layer for defining, decomposing and revising problems in the domain layer

The steps in the problem/design satisfaction relation are realised as incremental π-resolution states (Burrow and Woodbury, 1999) termed a partial satisfier and represented here as TFSPartialSatisfier. Partial Satisfiers represent initial, intermediate (partial) and fully resolved solutions of a given description. A TFSPartialSatisfier composes descriptions and feature structures as follows:

$$\text{TFSPartialSatisfier} \equiv \bullet\text{PartialSatisfier: Desc: FeatureStructure} \quad (3)$$

The label PSat is used as a shorthand term for representing the relationship between satisfiers and descriptions in the substrate. A PSat composes a collection of TFSPartialSatisfiers:

$$\text{PSat} \equiv \bullet\text{TFSPartialSatisfiers} \quad (4)$$

In the domain layer, the view of a solution is encapsulated in solution state objects, written as SState. As an object, a solution state composes both resolved designs and partial satisfiers. The constituents of a solution state are shown in Figure 7. To a designer, a solution state provides a view on a developing design.

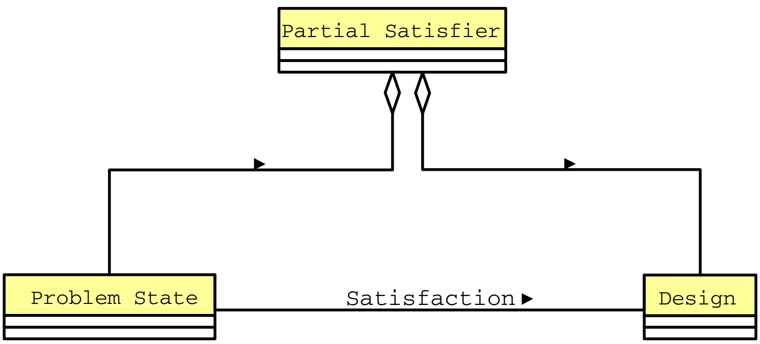


Fig. 7. Solution states compose partial designs with respect to an inheritance hierarchy of types and a partial satisfier

The realisation of the design as a typed feature structure and the satisfaction relation between problems and solutions as an incremental π-resolution state is shown in Figure 8. The solution state construct specifically adds intentionality concerning solutions to a problem specification. The connection of the solution state, SState to the formal substrate is given as follows:

$$\text{SState} \equiv \bullet\text{SState} \bullet \text{PState : Desc : TFSPartialSatisfier} \quad (5)$$

Summarising, the domain layer construct SState represents the notion of a design solution. In it, are embedded the symbol substrate concepts of description, the satisfaction of a description as satisfiers and the trace of intermediate solutions as partial satisfiers.

3.3 Choice

Problem formulation and generated solutions engender a large space of alternatives. These alternatives form a solution hierarchy and as designers revise solutions, a revision history of solutions is recorded. Support for the actions of the designer in making choices about solution alternatives and solution revision is necessary. Choice records commitments and thus the relation amongst functional requirements and characteristics of a physical structure that satisfy these requirements.

The connections between a problem and its possible solutions (partial or complete) are encapsulated as choices. Choices represent both the intentional commitments made by the designer (selection) and the alternative paths of resolution uncovered by the formalism. The choices made by the designer and the resultant choice points arising out of formal resolution are recorded as feature nodes in the domain layer. The FNode records user choices with respect to problem alternatives, incremental generation and the selection of solution alternatives. The feature node, FNode captures the relationship between a problem state, PState and an alternative design that is a partial solution to the problem, SState.

Further, a FNode composes typed feature structures in the underlying formalism. The user and the formalism participate in a process of incremental generation of partial solutions of a problem statement. First, in the selection of a FNode from the a collection of possible solutions. Second, in the specification of the next step of resolution. The connection between a FNode and a SState is given in the path form as follows:

$$\text{FNode} \equiv \bullet\, \text{FNode} \bullet \text{SState} : \text{FeatureStructure} \tag{6}$$

The FNode enables the user to make choices at a particular point in the problem formulation and solution generation process.

4 Conclusion

The paper presents an interaction model for supporting mixed-initiative in design exploration. It identifies the requirements of the domain layer and constructs a designer's view of exploration comprising problems, solutions, choices and history over the symbol level representation of design space exploration. The mapping is developed through the development of the concepts of: Problem state, Solution state, Choice and Satisfier space. In each of these constructs, the case for mixed-initiative is made through a three level exposition, the designer's view, the symbol system view and the mapping of the two through the domain layer constructs, PState, SState, and Choice. Problem states represent design problems. Solution states represent partial design solutions. Finally, the mapping of the domain layer to an interface construct, the FNode is demonstrated. Feature Nodes compose problem formulation and solution generation processes and support the exploration operations.

References

1. Burrow, A., Woodbury, R.: Design Spaces–The forgotten Artefact. In: Burry, M., Dawson, T., Rollo, J., Datta, S. (eds.) Hand, Eye, Mind, Digital, Mathematics and Design 2001, pp. 56–62. Deakin University, Australia (2001)
2. Burrow, A.: Computational Design and Formal Languages. Unpublished PhD Thesis, Department of Computer Science, The University of Adelaide (2003)

3. Burrow, A., Woodbury, R.: Π-resolution and Design Space Exploration. In: Augenbroe, G., Eastman, C. (eds.) Computers in Building: Proceedings of the CAADF 1999 Conference, pp. 291–308. Kluwer Academic Publishers, Dordrecht (1999)
4. Carpenter, B.: The Logic of Typed Feature Structures with applications to unification grammars, logic prorgams and constraint resolution. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge (1992)
5. Chien, S., Flemming, U.: Design Space Navigation: An Annotated Bibliography, Technical Report No EDRC 48-37-96, Engineering Design Research Center, Carnegie-Mellon University, Pittsburgh, PA, USA (1996)
6. Chien, S., Flemming, U.: Information Navigation in Generative Design Systems. In: Liu, Y., Tsou, J., Hou, J. (eds.) CAADRIA 1997, vol. 2, pp. 355–366. National Chia Tung University, Hsinchu (1997)
7. Cohen, R., Allaby, C., Cumbaa, C., Fitzgerald, M., Ho, K., Hui, B., Latulipe, C., Lu, F., Moussa, N., Pooley, D., Qian, A., Siddiqi, S.: What is Initiative? User Modeling and User-Adapted Interaction 8(3-4), 171–214 (1998)
8. Flemming, U., Woodbury, R.: Software Environment to Support Early Phases of Building Design (SEED): Overview. Journal of Architectural Engineering, ASCE 1(4), 147–153 (1995)
9. Guinn, C.I.: Mechanisms for Mixed-Initiative Human-Computer Collaborative Discourse. In: Joshi, A., Palmer, M. (eds.) Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics, pp. 278–285. Morgan Kaufmann Publishers, San Francisco (1996)
10. Rich, C., Sidner, C.L.: COLLAGEN: When Agents Collaborate with People. In: Johnson, W.L., Hayes-Roth, B. (eds.) Proceedings of the First International Conference on Autonomous Agents (Agents 1997), pp. 284–291. ACM Press, New York (1997)
11. Rich, C., Sidner, C.L.: COLLAGEN: A Collaboration Manager for Software Interface Agents. User Modeling and User-Adapted Interaction 8(3-4), 315–350 (1998)
12. Smith, R., Hipp, R.D.: Spoken Natural Language Dialog Systems: A Practical Approach. Oxford University Press, Oxford (1994)
13. Veloso, M.: Towards mixed-initiative rationale-supported planning. In: Tate, A. (ed.) Advanced Planning Technology, pp. 277–282. AAAI Press, Menlo Park (1996)
14. Veloso, M., Mulvehill, A., Cox, M.T.: Rationale-Supported Mixed-Initiative Case-Based Planning. In: Innovative Applications of Artificial Intelligence, Providence, RI, pp. 1072–1077 (1997)
15. Woodbury, R., Burrow, A., Datta, S., Chang, T.W.: Typed Feature Structures and Design Space Exploration. Artificial Intelligence in Design, Engineering and Manufacturing 13(4), 287–302 (1999), Special Issue on Generative Design Systems
16. Woodbury, R., Datta, S., Burrow, A.: Erasure in Design Space Exploration. In: Gero, J. (ed.) Artificial Intelligence in Design 2000, pp. 521–544. Kluwer Academic Publishers, Dordrecht (2000)

Intertwined Modeling and Implementation of Interactive Systems Using HOPS

Anke Dittmar and Peter Forbrig

University of Rostock, Germany
{anke.dittmar,peter.forbrig}@uni-rostock.de

Abstract. The paper investigates the co-evolution of models and implementations of interactive systems within the model-based design paradigm. A view of implementations as pure results of top-down model refinements is rejected. On the one hand, models inform or even drive further design and implementation steps. On the other hand, implementation ideas emerge during the iterative development process. They can be evaluated and further explored by models. In particular, selective modeling allows to focus attention on certain aspects of the interaction.

Higher-Order Processes Specifications (HOPS) describe interactive systems from different viewpoints and at different levels of granularity. The HOPS tool is used to suggest techniques for intertwining modeling and implementation activities. Object-oriented implementations in Java can be assigned to HOPS models. Their animation results in model-guided prototyping open for both empirical and analytical evaluation.

Keywords: Model-Based Design, Human-Centered Design, Viewpoints, Co-Evolution of Representations.

1 Introduction

Human-Centered Design (HCD) is based on the assumption that the development of interactive systems is rarely achievable without iterations and without an active participation of users and other stakeholders. An exclusive application of analytic techniques and design methods is rejected for several reasons. First, it is neither possible to fully understand existing working practices and to specify all requirements of the system under development in advance nor is it possible to foresee all effects of integrating new or modified artifacts into daily practices. Second, the design practice itself is considered as process of intertwined activities contributing to requirements analysis, prototyping, conceptual design and implementation. Evaluation steps guide this process.

This paper explores the co-development of models and implementations within the Model-Based Design (MBD) paradigm in order to support the ideas of HCD more effectively. Current model-based approaches mostly rely on the use of task models and domain models to derive specifications, prototypes or implementations for interactive devices [1]. The implementation is basically seen as a result of a top-down refinement. Typical techniques to produce lower-level descriptions from higher-level

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 194–203, 2011.
© Springer-Verlag Berlin Heidelberg 2011

descriptions are tool-supported mappings and transformations. In previous work, we have shown that current modeling practices tend to be specification-driven [2]. Task models are geared towards functional specifications or user interface specifications. As a consequence, they are often considered as complete and consistent descriptions but may fail to support some of the original objectives of task analysis such as the reflection of task allocation options or the selective consideration of usability problems. A co-evolution of implementations and models describing selected aspects of interactive systems may help to alleviate some of the above mentioned drawbacks of MBD practices. It may allow for easier integration with other analytic and empirical HCD techniques.

In this contribution, Higher-Order Processes Specifications (HOPS) are used to illustrate the idea of co-evolving models and implementations of interactive systems. HOPS is a universal specification formalism with high-level concepts for describing interaction from different viewpoints (e.g. [3]). It combines the description of behavioral and structural aspects of a system in the same notation. Tool support makes it possible to animate HOPS models and to map them to object-oriented implementations in Java at arbitrary levels of granularity. This allows a certain freedom in implementation activities, but also a selected control of implementations by models. Selective modeling helps to focus attention on specific aspects, for example on the identification of usability problems or on the exploration of design options [4].

The paper is organized as follows. Sect. 2 relates this work to existing design approaches. In Sect. 3, our general view on the co-evolution of models and code is given. Sect. 4 introduces an example scenario which is used throughout Sect. 5 to explore techniques for intertwining modeling and implementation activities by using HOPS. The paper closes with a short summary.

2 Related Work

In her critique on the waterfall model, Pfleeger points out: “Software is not developed like that; rather it evolves as the problem becomes understood and the alternatives are evaluated. Thus, software is a creation process, not a manufacturing process. The waterfall model tells us nothing about the typical back-and-forth activities that lead to creating a final product” [5]. Parnas and Clements state that “we will never see a software project that proceeds in the "rational" way…Many of the details only become known to us as we progress in the implementation. Some of the things that we learn invalidate our design and we must backtrack. Because we try to minimize lost work, the resulting design may be one that would not result from a rational design process” [6]. It is argued (for example in [7]) that requirements, specifications and implementations are intimately intertwined.

In comparison with other software development approaches, HCD emphasizes the central role of empirical and analytic evaluation activities in interactive systems design (Fig. 1a). A combination of different evaluation techniques throughout the design process is recommended to assess the three aspects of technology use: functionality, usability and the users’ experience of the interaction. The notion of task is at the heart of most of them in order to test systems or to reflect design options.

Examples are Cognitive Walkthrough [8] or think-aloud techniques. Although HCD practices have been refined over the years, the question of how to feed back results of evaluation steps into design in an effective way still remains.

Model-based design approaches also consider knowledge about user tasks as essential for developing interactive systems but otherwise rather rely on "classical" software engineering thinking. The Cameleon Reference Framework [9] describes the most popular method to exploit task models (Fig. 1b). Task and domain models serve as starting point for deriving final user interfaces via several levels of abstraction. Typical techniques are tool-supported mappings and transformations. MBD supports model reuse across different platforms and programming languages. It also supports a consistent development of interfaces which have to work in different contexts, e.g. in distributed environments. However, the use of task models is biased towards functional and user interface specifications (e.g. [10]).

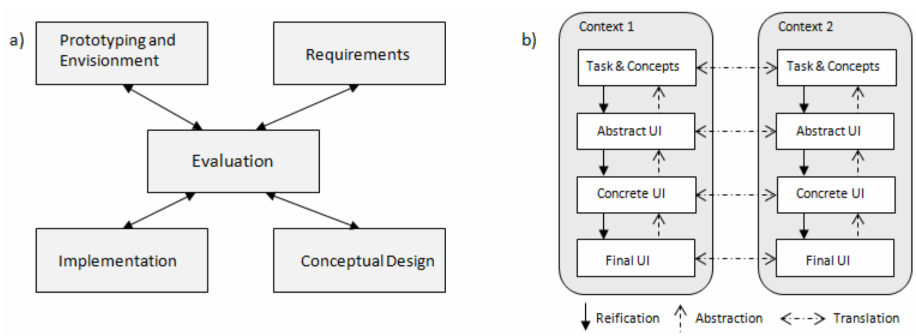


Fig. 1. a) Modified Star Life Cycle [11], b) Cameleon Reference Framework [9]

In the context of this paper, we are mainly concerned with improved techniques to co-evolve models and implementations in order to support more effective cycles of evaluation and design activities within the model-based paradigm.

3 Co-evolution of Models and Implementations

Task-based approaches in HCI have been criticized because of their narrow view on human activity. Alternative approaches as they are discussed e.g. in [12] put more emphasis on the interplay between actors and artifacts. It is assumed that available resources (external artifacts but also internal ones like plans) determine the actor's behaviors in actual situations. People experience their environment as activity spaces and change them according to their goals. The deeper their understanding of the environment the more fine-tuned their interventions.

If we apply this view on the model-based development process itself the overly dominant role of task models is questioned. Furthermore, the understanding of models as selective descriptions of phenomena from certain points of view is more emphasized. Task models describe how to act in the world to achieve a certain goal. Artifact models describe domain objects in terms of attributes and actions serving different purposes in different contexts. Device models, in particular, may describe possible usage scenarios of interactive devices in terms of visible information and enabled action sequences (dialogs). A prototype represents concrete human-computer interactions. It is represented by code in a programming language and is experienced in a different way than abstract models. Every representation can be considered as an artifact which evokes certain responses. Developers are responsible for a reasonable combination of representations to get a balanced view on the design problem.

As common in HCD approaches, design activities can happen in any order and are "glued together" by evaluation steps. The use of representations depends on the objectives of the actual step. For example, in order to explore task allocation options, models of tasks, artifacts and roles can be used as resources to develop appropriate device models. A formal specification can be evaluated by a corresponding prototype. Models can be used or firstly developed for the assessment of emerging implementations. Changes to these models then feed back into later design steps.

In this paper, we suggest selective mappings of models and code which make it possible to link the animation of models and the execution of prototypes. Such model-guided prototyping supports an application of analytic and empirical techniques to evaluate specific aspects of the system under design.

4 Example Scenario

Mastermind is a code breaking logic game for two players that is probably well known to most of the readers. Fig. 2 gives a glimpse on the problem to design an interactive version of this game. In the center of the figure, a task model is sketched out. It is annotated by the roles of the players. Numbered circles indicate relevant task domain concepts (code, guess, marking) and their reification in the depicted game versions. In our design scenario, the interactive version may be required to support different allocations of tasks. For example, both players are humans or the computer plays the role of the code maker. Throughout the rest of the paper, the example scenario is used to explore the above mentioned ideas in a concrete setting.

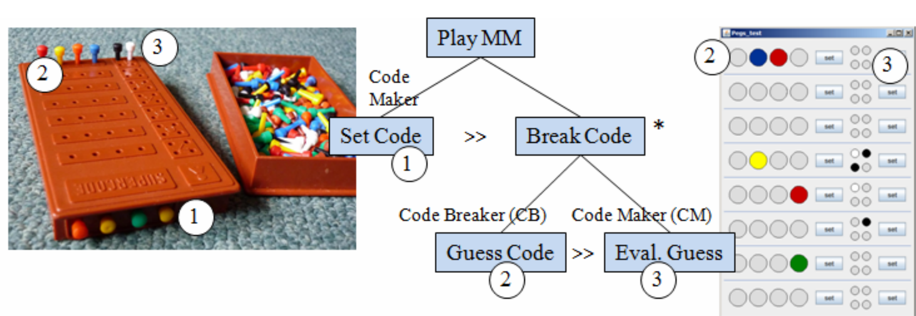


Fig. 2. Views on playing Mastermind

5 Co-evolution of Models and Code with HOPS

HOPS (Higher-Order Processes Specifications) is a formal specification formalism with tool support to describe interactive systems. For reasons of brevity, a short and informal introduction to the formalism is given only. A more detailed explanation is to be found e.g. in [2,3]. In the context of HOPS, systems are considered as compositions of interacting sub-systems. They are specified by *processes*. The structure of a process is determined by components, operations and sub-processes. *Operations* refer to the smallest units of behavior that are of interest in the actual modeling context. The behavior of a process is defined by a set of sequences of operations. *Sub-processes* refer to partial behavior of a process and can be specified by partial equations and the use of temporal and structural operators. They can be applied for several purposes, for example for describing certain states of components or for creating structures such as hierarchies. *Components* as instances of previously defined processes serve to model sub-systems. Processes without components are called *basic processes*. Their operations are defined by names only. Processes with components are called *higher-order processes*. They can define new atomic behavioral units at a higher level of abstraction by “melting” sequences of components’ operations. A higher-order process can use its sub-processes and higher level operations to partially control the otherwise autonomous behavior of its components.

5.1 Basic Processes in HOPS

Fig. 3 shows in the bottom right section an example of a basic process to describe a peg space. Three operations are defined to model that a space can get a peg, the peg can be removed and the state of the space can be declared as fixed (lines 4-6). This view on a peg space may be appropriate in the context of the mastermind game. In HOPS,

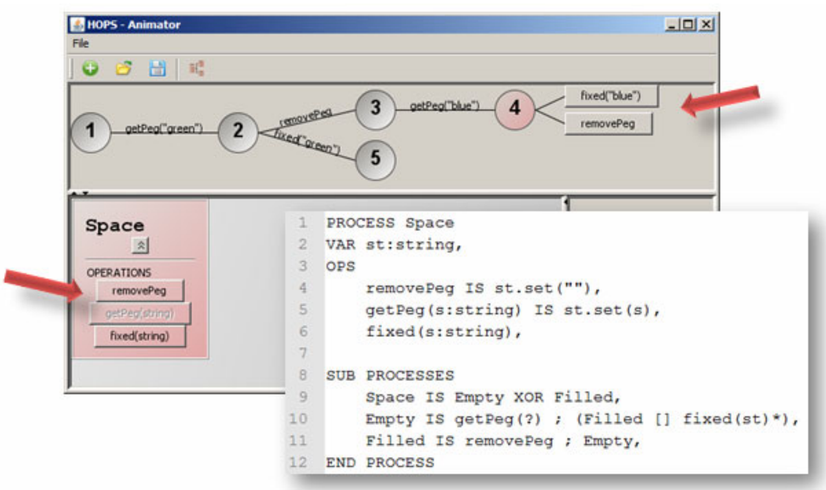


Fig. 3. Definition of process *Space* and animation run of a process instance

partial equations (lines 9-11) define temporal constraints on operation sequences by using temporal operators as known, for example, from CCT [1].

Temporal operators in HOPS:				
	;	sequence	#n	n-fold iteration
	[]	alternative	[…]	option
	|||	concurrency	|>	disabling
	*/+	iteration	[>	interruption

The HOPS tool offers interactive model animation. At each step of an animation run, the user can choose to execute one of the actual enabled HOPS operations. In Fig. 3, enabled operations are presented and can be activated in two ways (marked with red arrows): within the actual animation tree in the top of the figure and within the visualization of the process instances the HOPS model consists of. Each path of the animation tree represents the prefix of a valid operation sequence of the model. The user can "jump" between nodes of the tree to explore specified behavior.

5.2 Higher-Order Processes

Higher-order processes in HOPS contain components that are themselves process instances. Hence, HOPS models can be considered as hierarchical structures of process instances (also called component trees)[1]. Higher-order processes not only allow to describe the coordination between the components but also to conflate sequences of components' operations into new atomic behavioral units (operations). This helps to increase the level of abstraction in a description. Often, a higher-order process focuses on some behavioral aspects of its components only. Those lower-level operations which are not known to and controlled by the higher-order process are still observable units of behavior and can be executed during model animations. They only obey the constraints already specified in components. In other words, the behavior of a higher-order process P is defined by a set of sequences of operations which are either defined by P itself or which are to be found in behavior descriptions of components but not controlled by P.

An example is given in Fig. 4. Process *Guess_code* consists of three components: *s1* and *s2* are instances of *Space* and behave like sub-process *Empty* (see Sect. 5.1); *p* is an instance of *Pegs* which models colored pegs that can be taken arbitrary often.

```
PROCESS Pegs
OPS
    take(s: string),
SUB PROCESSES
    Pegs IS (take("yellow") [] take("red"))*,
END PROCESS
```

Process *Guess_code* serves as a task description of making a guess in the context of the mastermind game. In this case, two code pegs are to be placed. Two higher level operations are defined: *setPeg* takes a peg of a certain color and places it either at s1 or s2 (lines 7-8). Operation *fixed* declares the guess as completed. It conflates sequence *s1.fixed(c1) ; s2.fixed(c2)* into a new unit (lines 9-10). The higher-order process focuses on the placing of pegs only. The control of operation *removePeg* remains in the *Space* components. This is also to be seen in the component tree in Fig. 4. The button

[1] Component sharing results in directed acyclic structures though.

representing *removePeg* is enabled in the components' nodes. At the depicted stage of the animation, the guess could be finished because both spaces are filled and so operation *fixed("yellow","yellow")* is enabled or the peg could be removed from space s1 or s2 (*s1.removePeg/ s2.removePeg*).

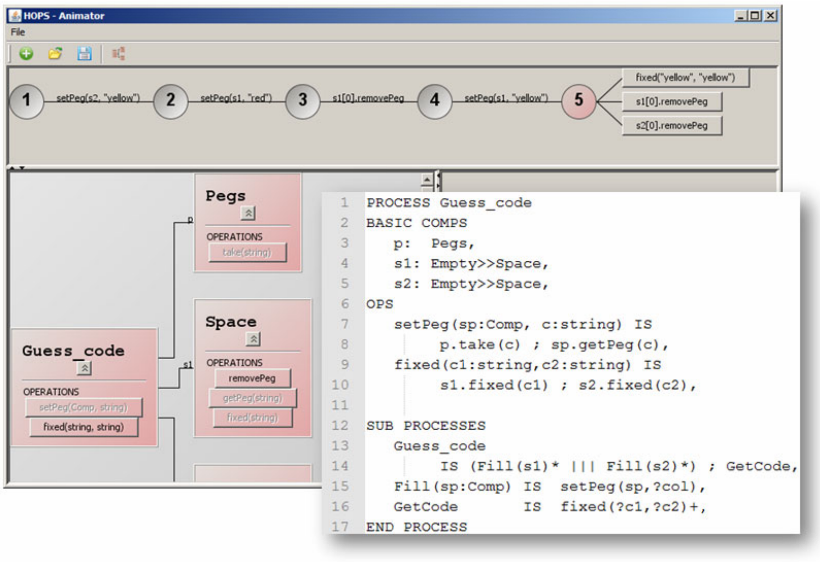


Fig. 4. A higher-order process with three components and screenshot of animation run *⟨setPeg(s2,"yellow"),setPeg(s1,"red"),s1.removePeg,setPeg(s1,"yellow")⟩* with the component tree in bottom left section.

The HOPS notation makes it possible to describe a system from different viewpoints and to integrate them by using higher-order processes. It supports both bottom-up as well as top-down thinking. On the one hand, the behavior of a process is determined by its components. On the other hand, it has partial control over them and can define new behavioral units. Partly redundant models can be used to enable description of emerging constraints and of distributed control. In the example, process *Guess_code* reflects a task perspective. It is said that the goal (to guess the code) can be achieved by sequences such as *⟨setPeg(..),..,setPeg(..),fixed(..)⟩*. The components in this model describe the domain in terms of artifacts. Although they are related to the task, they still reflect a wider range of possible actions (e.g. *removePeg*, see Fig. 4). In this way it is possible to describe a more liberal task-oriented behavior.

5.3 Mappings between Models and Implementations

The HOPS tool enables an automated mapping of HOPS process instances to Java objects. Additionally, Java method calls can be assigned to HOPS operations in process definitions. If an operation is executed during model animation the corresponding Java method calls are executed as well.

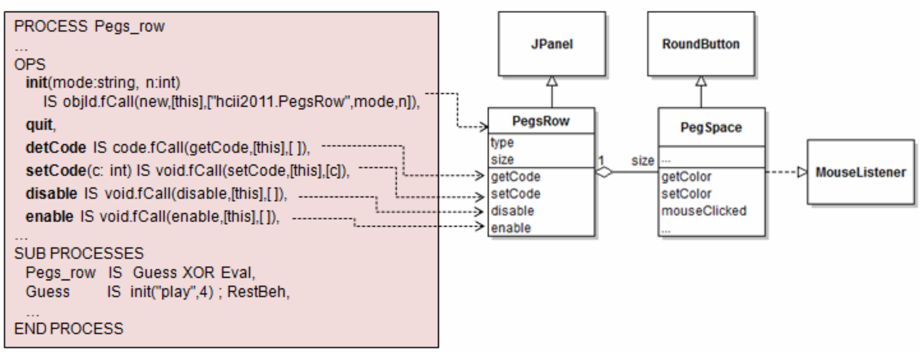


Fig. 5. HOPS process describing some actions on rows of code pegs or key pegs

Fig. 5 and 6 illustrate how this means can be exploited for selective modeling and control of an example implementation of peg rows as it is needed for creating interactive mastermind boards. Parts of HOPS process *Pegs_row* are shown on the left side of Fig. 5. The class diagram on the right side sketches out the structure of the Java implementation. Dotted arrows indicate mappings from HOPS operations to Java methods. As to be seen, the HOPS model considers overall behavior of peg rows such as their enabling and disabling but ignores details such as the setting of single pegs.

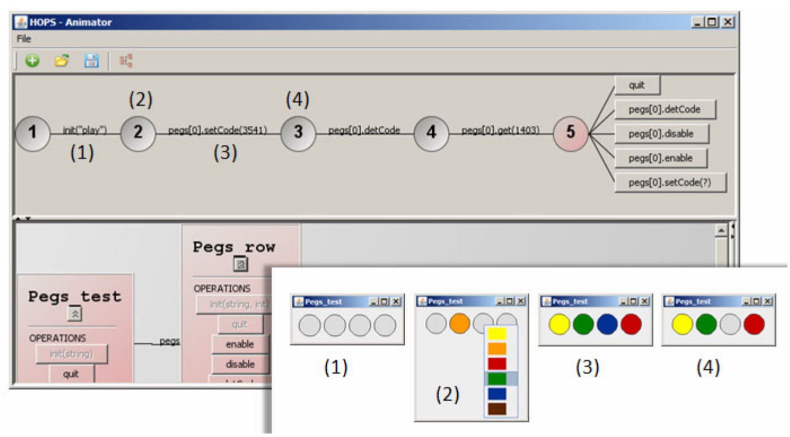


Fig. 6. The animation of process *Pegs_row* (embedded in a test process) selectively controls the assigned Java implementation (bottom right)

Fig. 6 shows the HOPS animator and a sequence of screenshots of the assigned peg row prototype. The execution of HOPS operation *init("play")* creates a row of code pegs (1). The effect of HOPS operation *setCode(3541)* on the prototype is marked by (3). However, users can also interact with the implementation in ways which are not considered by the HOPS model. They can set single pegs (2) and remove pegs (4).

Process *Pegs_row* is a device model. It could already be used to rethink the behavior of the implementation depicted in Fig. 2 where pegs can be placed arbitrarily on the board. The next section shows how this device model is combined with models of tasks and actors in order to reflect design options for supporting different task allocations of the example system (as required in Sect. 4).

5.4 Co-development of Representations in the Example Scenario

The focus of the example co-development is on the exploration of system behavior that supports specific allocations of tasks. In the example, the computer as well as users can play both the role of code breakers and of code makers. A sharing of tasks is also possible but not considered here for the sake of simplicity.

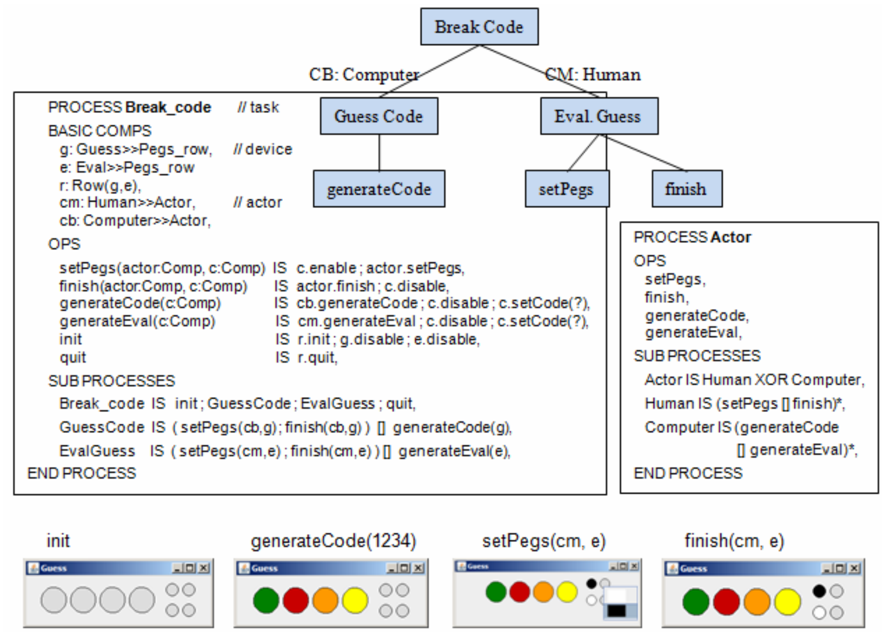


Fig. 7. Task model *Break_code* as a composition of device models and actor models (top part) partly controls the behavior of the assigned implementation (bottom part)

Fig. 7 shows a refined part of the task model in Fig. 2. In addition, task *Break_code* is described as interaction of HOPS components modeling the device and two actors. In the figure, the code maker is a user and the system acts as code breaker. The *Actor* process reveals task-related differences in the behavior of users and the system. In particular, users explicitly complete a guess or evaluation (operation *finish*). The task process mainly describes strategies to enable and disable peg rows dependent on the actual function allocation. The bottom part of Fig. 7 shows an animation run in which the computer guesses the code and then the user can place key pegs to evaluate the guess. This model-guided prototyping allows reasoning over selective conceptual models but also the experience of interacting with the "real" implementation.

6 Summary

The paper has demonstrated how the HOPS formalism and corresponding tool support can be applied to a co-evolution of code and models representing different viewpoints of an interactive system. Design processes are seen as guided by different design representations rather than controlled, for example, by task models. The modeling of selected aspects of interaction and model-guided prototyping enable a deeper integration of analytic and empirical evaluation steps into design activities. This should be investigated in further research. The presented work can be seen as continuation of previous work on refining model-based design practices [2, 3].

References

1. Paternó, F.: Model-Based Design and Evaluation of Interactive Applications. Springer, Heidelberg (2000)
2. Dittmar, A., Forbrig, P.: Task-based design revisited. In: Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2009), pp. 111–116. ACM, New York (2009)
3. Dittmar, A., Harrison, M.D.: Representations for an iterative resource-based design approach. In: Proceedings of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems(EICS 2010), pp. 135–144. ACM, New York (2010)
4. Bellotti, V., Shum, S.B., MacLean, A., Hammond, N.: Multidisciplinary modelling in HCI design...in theory and in practice. In: Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1995 (1995)
5. Pfleeger, S.L.: Software Engineering: Theory and Practice, 2nd edn. Prentice-Hall, Englewood Cliffs (2001)
6. Parnas, D.L., Clements, P.C.: A rational design process: How and why to fake it. IEEE Trans. Softw. Eng. 12(2), 251–257 (1986)
7. Siddiqi, J.: Challenging Universal Truths of Requirements Engineering. IEEE Software 11(2), 18–19 (1994)
8. Polson, P.G., Lewis, C., Rieman, J., Wharton, C.: Cognitive Walkthroughs: A Method for Theory-Based Evaluation of User Interfaces. International Journal og Man-Machine Studies 36(5), 741–773 (1992)
9. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
10. Luyten, K., Clerckx, T., Coninx, K., Vanderdonckt, J.: Derivation of a Dialog Model from a Task Model by Activity Chain Extraction. In: Jorge, J.A., Jardim Nunes, N., Falcão e Cunha, J. (eds.) DSVIS 2003. LNCS, vol. 2844, pp. 203–217. Springer, Heidelberg (2003)
11. Benyon, D., Turner, P., Turner, S.: Designing interactive systems: people, activities, contexts, technologies. Addison-Wesley, Reading (2005)
12. Wright, P.C., Fields, R.E., Harrison, M.D.: Analyzing human-computer interaction as distributed cognition: the resources model. Human Computer Interaction 15(1), 1–42 (2000)

HCI Patterns as a Means to Transform Interactive User Interfaces to Diverse Contexts of Use

Jürgen Engel[1], Christian Märtin[1], and Peter Forbrig[2]

[1] Augsburg University of Applied Sciences, Faculty of Computer Science, Friedberger Strasse 2a, 86161 Augsburg, Germany
[2] University of Rostock, Institute of Computer Science, Albert-Einstein-Strasse 21, 18059 Rostock, Germany
{Juergen.Engel,Christian.Maertin}@hs-augsburg.de, {Peter.Forbrig}@uni-rostock.de

Abstract. This paper introduces a pattern-based method for transformation of user interfaces of interactive applications to diverse contexts of use. The method is demonstrated with the help of various examples taken from existing software solutions. The related pattern transformation rules are derived from the samples and in turn expressed in a pattern format, the so-called transformation patterns.

Keywords: HCI patterns, user interface, pattern-based context transformation, transformation patterns, interactive systems, PLML.

1 Introduction

In the past electronic appliances and devices, e.g. telecommunication devices, usually were comprised of proprietary hardware. Nowadays such products are increasingly implemented on a software basis, e.g. as embedded applications, web clients or Apps. These types of interactive software systems steadily become more important.

As in most application domains users can choose from a vast variety of competing products usability and user interface aspects become substantial. But user interface development in particular absorbs a significant portion of the total development effort [15]. Emerging platforms, such as smart phones, other mobile devices, and new web technologies are the driving forces for even higher UI complexity.

Due to these facts we focus our research on options for automation in the UI development process. Here a major aspect is pattern- and model-based UI modeling and UI code generation. Within this paper we would like to share the findings of our ongoing work on transforming user interfaces to diverse contexts of use.

2 Related Work

Patterns and pattern languages have a growing impact on the disciplines of HCI and web engineering. Christopher Alexander introduced patterns for solving problems in architecture and urban planning. He defined a pattern as a three-part rule which expresses a relation between a certain context, a problem, and a solution [1]. In the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 204–213, 2011.
© Springer-Verlag Berlin Heidelberg 2011

1990s patterns were adopted by software engineers and software architects for promoting the reuse of high-quality design solutions [8]. For over a decade now patterns and pattern languages have also entered the fields of HCI, usability engineering [12], user experience [18] and organizational workflow [9].

Over the years various HCI pattern catalogues have been developed and published providing valuable and reusable design knowledge. Examples are Jenifer Tidwell's *Designing Interfaces, Patterns for Effective Interaction Design* [17], Martijn van Welie's Patterns in Interaction Design [19], Todd Coram's and Jim Lee's *A Pattern Language for User Interface Design* [3], or the community-driven UX and UI pattern library *Quince* operated by Infragistics [10].

However, most of the available pattern collections lack an appropriate organizational structure in order to facilitate pattern selection and ensure the overall coverage of domain dependent and independent modeling and design problems. Manageability aspects of various existing UI pattern catalogues are discussed and compared in [4]. In [13] a structured approach both for designing hierarchically organized HCI pattern languages and controlling the selection of the really needed patterns during the software development process is introduced.

Another significant shortcoming in the area of patterns is that the various authors usually describe their patterns in different and inconsistent styles. This makes it hard or even impossible to search, select and reference patterns across pattern collections. In a workshop which has been held during the CHI 2003 conference the participants aimed for unification of pattern descriptions and guidance for the authors. Hence the Pattern Language Markup Language (PLML) has been constituted. The current version PLML v1.1 stipulates that the documentation of a certain pattern should consist of the following elements: a pattern identifier, name, alias, illustration, descriptions of the respective problem, context and solution, forces, synopsis, diagram, evidence, confidence, literature, implementation, related patterns, pattern links, and management information [7].

The use of patterns can support software engineering activities in all phases of the software development life-cycle. A comprehensive process for user interface engineering using patterns is e.g. introduced in [16].

In [11] it is demonstrated that HCI patterns can be used to migrate existing user interfaces of web applications across platforms considering the screen size of the respective user devices and the complexity of the data architecture. Two different approaches can be used for this type of migration: redesign and reengineering. While redesign is a simplified procedure for migrating directly from one to another platform-specific user interface, reengineering consists of an intermediate step of creating an abstract platform-independent UI model. In a case study redesign with navigation patterns is illustrated.

The current state-of-the-art in the growing field of HCI patterns and pattern-based software development approaches is documented in [2].

3 Pattern-Based Context Transformation

In [5] we introduced a pattern- and model-based framework for partially automated generation of UI code. One of its key components is a pattern repository containing the pattern definitions according to PLML v1.1 [7]. The patterns are hierarchically organized according to their respective levels of abstraction as outlined in [13]. On a top level view, the related taxonomy distinguishes patterns for navigation, content presentation and user interaction issues. The framework supports its users designing an abstract user interface model which serves as the basis for transformation into context-specific semi-abstract UI models. Finally the code of the target user interface is automatically generated from these semi-abstract models.

The UI model construction and transformation processes are based on diverse models, including task model, user model, device model, environment model, and data architecture model.

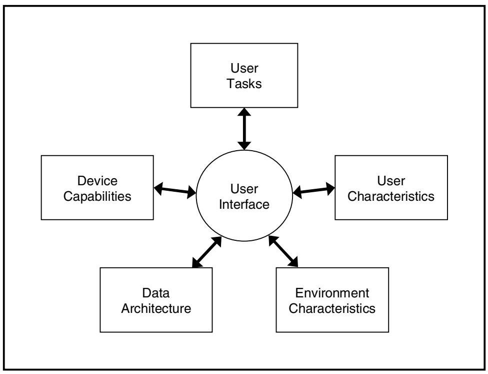


Fig. 1. User interface influence factors

The task model commonly incorporates the user activities, their types, hierarchical structure and temporal relationships. Preferably task models are represented as Concur Task Trees (CTT) [15]. In the user model we hold information about the user's preferences, expertise, possible disabilities and grade of distraction, for instance when the user is in charge of driving a car while interacting with a mobile device, e.g. a phone or navigation system. The device model holds data about input and output capabilities, screen size and resolution, bandwidth, storage capacity, processor power, and available widget sets. If not explicitly modeled, we use some of these values to determine the device's media capabilities, such as text, graphics, sounds, and audio and video streaming. The environment model may indicate possible problems due to light irradiation, noise, or pollutant. Especially for web applications the data architecture model contains information about the complexity of the relevant content to be provided to the user [6], [11]. From these models one can construct various contexts of use during design time.

The starting point of the pattern-based UI model transformation process is a set of individual HCI patterns. Table 1 provides an overview of the patterns we had in focus during our analysis work even though the list is not depletive. For lack of space it is not possible to provide here the full-blown pattern definitions according to PLML

Table 1. Alphabetical list of HCI patterns used for Transformations

ID	HCI Pattern Name	Brief Description
01	Accordion	Stack panels vertically or horizontally and open up one panel at the time while collapsing the other panels [19]
02	Alphabetical Index	Index contains links to pages providing details of the indexed terms
03	Bread Crumbs	Hierarchical navigation path from top level to current page making each step clickable [19]
04	Center Stage	Put the most important part into the largest subsection of the page and cluster secondary content around it [17]
05	Collapsible Panels	Create panels that can be opened or closed independently of each other [19]
06	Contextual Horizontal Menu	A horizontal menu displayed temporarily depending on a certain context
07	Contextual Vertical Menu	A vertical menu displayed temporarily depending on a certain context
08	Directory Navigation	Grouping links for items of the second hierarchy level under headings of the first level [19]
09	Drop-down Menu	A list of items appears when invoking the menu and one of the displayed items can be selected
10	Fly-out Menu	Combination of horizontal menu with sub-menu flying out while hovering over the main menu-item [19]
11	Portal Site	One overall home-page leading to several sub-sites [19]
12	Keyword Search	Offer a simple search capability to the user for locating required content by specifying keywords
13	List Builder	Present the total list and provide editing functionality next to it [19]
14	List View	Show a simple list of items
15	Most Relevant First	Put the most important information in the first place of a container element, e.g. a list of items
16	One-Window Drilldown	Show each page within a single window and as a user drills down replace the window contents completely with the new page [17]
17	Permanent Horizontal Menu	An always visible single-row menu bar providing a consistent set of links leading to key sections
18	Permanent Vertical Menu	An always visible vertical menu providing a consistent navigation tool across all pages
19	Repeated Menu	Repeat the main navigation on the bottom of the page in order top minimize scrolling effort [19]
20	Tab Menu	Put groups of content onto separate panels which can be accessed via tabs
21	Teaser Menu	Show a partial menu with a capability to expand to the full-blown menu, e.g. using a “show more” link [19]
22	Tiled Sections	Define separate sections of content and lay them all out on the page together [17]
23	Top Link	Provide a link to the top of the page at locations in the main content [19]
24	Two-Panel Selector	Put two panels side by side – one displaying a set of selectable items and the other showing the details [17]
25	Wizard	Lead the user through the interface step by step, doing tasks in a prescribed order [17]

v1.1, but the pattern names and a brief description. For these patterns we have defined reasonable transformation directives for specific contexts of use. These rules in turn can themselves be regarded as type of transformation or mapping patterns. They consist of descriptions of a problem, i.e. the transformation problem, a context, i.e. the modeled contexts of use, and a solution, i.e. the transformation directive. Thus we have a unified method to describe both, the underlying HCI patterns and the transformation rules.

The pattern-based UI model transformation is depicted with the help of the following examples taken from the analysis of existing software solutions. Here it is intended to provide user interfaces of an individual application for different devices, i.e. desktop PC and smart phone with a limited screen size.

Figure 2 shows in its upper part the desktop PC version of the world-wide-web homepage of the German public TV station ARD [20]. For instance, the main navigation of the web page is implemented according to the "Permanent Horizontal Menu" pattern (pattern ID 17 in table 1) and consists of 12 menu entries. For the small device version of the UI this pattern is initially transformed to the same pattern, but the resulting horizontal menu bar is truncated to only one single menu item "Navigation" as shown within the left screenshot in the lower area of figure 2. When clicking on this menu item a new screen is opened up according to the "One Window Drilldown" pattern (ID 16), showing the original full amount of menu options, but now according to the "Permanent Vertical Menu" pattern (ID 18).

Fig. 2. Desktop PC and Mobile Version of the ARD Internet Homepage

There exist different types of pattern-based transformations. At first there is the option to use the same pattern in the source as well as in the target user interface, i.e. the sameness on the abstract pattern-level. A second possibility is to apply the same patterns, but to vary in extensiveness, e.g. to include more or less items in a menu bar or to incorporate much or sparse content into a screen. The third option is to completely reorganize the structure and replace patterns by one or multiple different ones. Basically multiple transformation patterns for a specific pattern and a specific context may coexist, i.e. that there are transformation alternatives.

On basis this circumstance it is possible to derive a first transformation pattern definition as presented in table 2. Again, for lack of space, we merely included the transformation pattern attributes ID, problem, context, solution and illustration.

Table 2. Definition of Transformation Pattern #1

Attribute	Description
ID	T-001
Problem	How to transform a "Permanent Horizontal Menu" pattern (ID 17) under the given context?

Context	• Source: desktop PC user interface
	  • Display size: 15 in. or bigger
	  • Display resolution: 1024x768 or better
	• Target: smartphone device UI
	  • Display size: 35x41 mm or alike
	  • Display resolution: 176x208 px or alike

Solution	1. Apply a "Permanent Horizontal Menu" pattern (ID 17), but drastically reduce the amount of items, e.g. only one item named "navigation" 2. Apply a "One-Window Drilldown" patter (ID 16) in order to open up a new screen 3. Apply a "Permanent Vertical Menu" pattern (ID 18) including the original list of items.
Illustration	Please refer to figure 2.

The pattern-based transformation is not limited to navigation patterns, but can also be applied for content presentation and interaction patterns. Figure 3 shows in its upper part the main page of the internet email service of t-online.de, an online service of the German telecommunications carrier Deutsche Telekom AG [21]. This page is implemented according to the "Portal Site" pattern (ID 11). Amongst others it contains a "Two-panel Selector" pattern (ID 24) as indicated by the two ellipses in the screenshot. The left part allows for selecting an individual type of email messages, while the other provides a list of emails of the chosen type.

The fundamental email type selection function is transformed to a "List View" pattern (ID 14) implemented as entry page for the mobile UI. When clicking on one of the displayed list items a subsequent page is opened up according to the "One-Window Drilldown" pattern (ID 16) showing the realization of a "List Builder" pattern (ID 13) providing the list of respective emails. The particular list entries are arranged as multi-line items so that the most relevant information can be displayed on a small device. In summary the "Two-panel Selector" pattern of the large screen is replaced by several different patterns within the mobile UI.

The related transformation pattern definition reads as shown in table 3.

Fig. 3. Desktop PC and Mobile Version of the Email Service t-online.de

Table 3. Definition of Transformation Pattern #2

Attribute	Description
ID	T-002
Problem	How to transform a "Two-Panel Selector" pattern (ID 24) under the given context?
Context	Transformation from desktop PC to smartphone UI as contemplated in pattern T-001
Solution	1. Apply a "List View" pattern (ID 14) incorporating the list of selectable items of the overview panel. 2. Apply a "One-Window Drilldown" patter (ID 16) in order to open up a new screen 3. Apply a "List Builder" pattern (ID 13) for showing the respective details.
Illustration	Please refer to figure 3.

Another example is the German weather service wetter.de [22] where users can request weather forecast information for individual regions or cities. Both the PC and the mobile version identically request the related user input by respectively applying the "Keyword Search" pattern (ID 12). Figure 4 illustrates on its right hand side the search results for the city of Augsburg (Germany) in the desktop PC manner.

On the left side it is shown the respective output as it appears on the mobile device. For demonstration purposes the entire virtual mobile page is displayed while the black rectangle indicates the visible part on the physical screen. The user has the option to scroll up and down in order to see the whole stuff.

On the desktop PC page the content is provided in a "Tab Menu" pattern (ID 20) style for different weather forecast periods. Each panel is organized according to the "Tiled Sections" pattern (ID 22). The columns of the 3-day forecast tab contain the relevant information for today, tomorrow and the day after tomorrow. In the upper

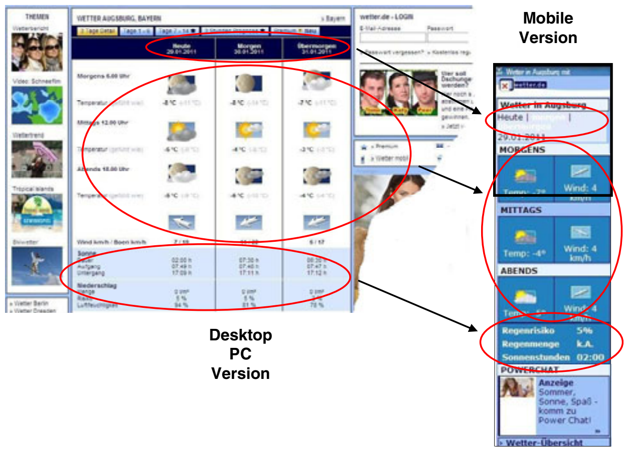


Fig. 4. Desktop PC and Mobile Version of the Weather Service wetter.de

rows of the matrix you see icons and temperature values for morning, noon and evening. The lower rows provide additional daily details about sunshine duration and precipitation.

Within the mobile version the different tabs have no equivalent because solely a forecast period of three days is available. The column headings of the "Tiled Sections" pattern are represented according to the "Permanent Horizontal Menu" pattern (ID 17). Though the horizontal menu disappears when scrolling down, it is located on each of the result pages and therefore is a permanent menu. The rows for morning, noon and evening values are translated into a type of "List View" pattern (ID 14) appearance. A selection of the daily detail measures are also displayed on the basis of the "List View" pattern.

Based on this actual situation the following three transformation patterns can be defined.

Table 4. Definition of Transformation Pattern #3

Attribute	Description
ID	T-003
Problem	How to transform a "Keyword Search" pattern (ID 12) under the given context?
Context	Transformation as contemplated in pattern T-001
Solution	Just apply a "Keyword Search" pattern (ID 12)
Illustration	-

In a similar manner it is also possible to transform specific user interfaces for other contexts of use, e.g. to create different UI for novice and expert users. Another idea is to adapt UIs for visually handicapped (not blind) people where the UI works in general similar to a UI for small screen sizes, but the elements and artifacts are displayed on a large screen in a magnified manner.

Table 5. Definition of Transformation Pattern #4

Attribute	Description
ID	T-004
Problem	How to transform a "Tab Menu" pattern (ID 20) under the given context?
Context	Transformation from desktop PC to smartphone UI as contemplated in pattern T-001
Solution	Pick the most important tabbed panel and apply a "One-Window Drilldown" pattern (ID 16) for arranging the related content onto the resulting new page. Let all other tabbed panels unconsidered.
Illustration	Please refer to figure 4

Table 6. Definition of Transformation Pattern #5

Attribute	Description
ID	T-005
Problem	How to transform a "Tiled Sections" pattern (ID 22) under the given context?
Context	Transformation from desktop PC to smartphone UI as contemplated in pattern T-001
Solution	1. Apply a "Permanent Horizontal Menu" pattern (ID 17) and use the column headings of the "Tiled Section" pattern as menu items. 2. Apply a "List View" pattern (ID 14) and incorporate the content pieces from the tiled sections. If applicable, define groups of contents and apply "List View" patterns for each of the groups.
Illustration	Please refer to figure 4

4 Conclusion

In this paper we have described a pattern-based method to transform user interfaces of interactive systems to diverse contexts of use. The main focus of our current work lies on modeling necessary relationships between the various patterns and defining a comprehensive set of transformation patterns, i.e. the rules required for pattern-driven transformation of existing abstract and semi-abstract UI models to concrete context-related target user interfaces. In a subsequent step we will put our emphasis on automation issues of the transformation process.

References

1. Alexander, C., Ishikawa, S., Silverstein, M.: A pattern language. Oxford University Press, Oxford (1977)
2. Breiner, K., et al. (eds.): Proceedings of the 1st International Workshop on Pattern-Driven Engineering of Interactive Computing Systems – PEICS 2010, Berlin, Germany, pp. 24–27 (2010)

3. Coram, T., Lee, J.: A Pattern Language for User Interface Design, http://www.maplefish.com/todd/papers/experience.html
4. Deng, J., Kemp, E., Todd, E.G.: Managing UI Pattern Collections. In: Proc. CHINZ 2005, Auckland, New Zealand, pp. 31–38. ACM Press, New York (2005)
5. Engel, J., Märtin, C.: PaMGIS: A Framework for Pattern-based Modeling and Generation of Interactive Systems. In: Proceedings of HCI International 2009, San Diego, USA, pp. 826–835 (2009)
6. Engelberg, D., Seffah, A. (eds.): Design Patterns for the Navigation of Large Information Architectures. Proceedings of 11thAnnual Usability Professional Association Conference, Orlando, Florida, July 9-12 (2002)
7. Fincher, S., et al.: Perspectives on HCI Patterns: Concepts and Tools (Introducing PLML). In: CHI 2003 Workshop Report (2003)
8. Gamma, E., Helm, R., Johnson, R., Vlissides, J. (eds.): Design Patterns: Elements of Reusable Object Oriented Software. Addison-Wesley, Reading (1995)
9. Guerrero Garcia, J., Vanderdonckt, J., Gonzalez Calleros, J.M., Winckler, M.: Towards a Library of Workflow User Interface Patterns. In: Graham, T.C.N. (ed.) DSVIS 2008. LNCS, vol. 5136, pp. 96–101. Springer, Heidelberg (2008)
10. Infragistics, Inc., Quince, http://quince.infragistics.com
11. Javahery, H., Seffah, A., Engelberg, D., Sinnig, D.: Migrating User Interfaces Across Platforms Using HCI Patterns. In: Multiple User Interfaces, pp. 241–259. Wiley & Sons, Chichester (2004)
12. Marcus, A.: Patterns within Patterns. Interactions 11(2), 28–34 (2004)
13. Märtin, C., Roski, A.: Structurally Supported Design of HCI Pattern Languages. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4550, pp. 1159–1167. Springer, Heidelberg (2007)
14. Myers, B.: Why are Human-Computer Interfaces Difficult to Design and Implement? Carnegie Mellon School of Computer Science Technical Report, no. CMU-CS-93-183 (1993)
15. Paternò, F.: Model-based Design and Evaluation of Interactive Applications. Springer, London (2000)
16. Seffah, A., Gaffar, A.: Model-based User Interface Engineering with Design Patterns. J. of Systems and Software 80(8), 1408–1422 (2007)
17. Tidwell, J.: Designing Interfaces, Patterns for Effective Interaction Design, http://www.designinginterfaces.com
18. Tiedtke, T., Krach, T., Märtin, C.: Multi-Level Patterns for the Planes of User Experience. In: Proc. of HCI International, Las Vegas, Nevada USA, July 22-27. Theories Models and Processes in HCI, vol. 4, Lawrence Erlbaum Associates, Mahwah (2005)
19. van Welie, M.: Patterns in Interaction Design, http://www.welie.com
20. Homepage of the German Public TV Station ARD, http://www.ard.de
21. Homepage of t-online, an Online Service of Deutsche Telekom AG, http://www.t-online.de
22. Hompage of a German weather forecast portal, http://www.wetter.de

Process Choreography for Human Interaction Computer-Aided Simulation

Carlos Fernandez-Llatas, Juan Bautista Mocholí,
Pilar Sala, and Juan Carlos Naranjo

TSB-ITACA Universidad Politecnica de Valencia, Spain

Abstract. Design errors can suppose a unaffordable load for the production costs. Allowing product designers to de ne behavioral patterns that describe the interaction of future users of the system can reduce the number of design errors. These patterns can be used to simulate how users respond to stimuli of products detecting problems at early stages of product development. Choreography systems to simulate the interaction among devices and services defined using commercially available workflow engines have been used in previous work (as the European project VAALID). However, the complexity of human behavior models requires much more expressive workflow languages for their definition.

In this work, a highly expressive Workflow engine is presented. This system solves the problem of expressiveness in the representation of the interaction of human behavior models in the VAALID project.

1 Introduction

Currently, the hard competency among enterprises requires that the designed products ts perfectly with the customers needs. In this way, the design errors suppose large cost overruns. For that, the early detection of problems in the product creation cycle of life can suppose significant economical savings to companies. This problem worsens when the products target special needs persons like people with age-related impairments or people with disabilities. This is be-cause those users have additional constraints and usually the designers are not well aware of their implications and don't take them in full consideration during the creation of prototypes [2]. In those cases, accessibility and usability, as well as the seven principles of the Universal Design [9] or Design for All are crucial concepts, but still not fully incorporated in the mainstream industry, being one of the main reasons for this the lack of tools to support the designers to adopt these methods [2]. The development of simulation techniques and tools to help the designers to create and evaluate new systems and objects has been a hot topic in the last years because simulating the interaction between users and products during the design phase allows to detect conceptual errors in early stages of the development process.

In order to simulate the interaction among the system components in design phase it is necessary to describe the components execution flow in a formal way. This is commonly done by using workflow technology [11]. Workflow technology allows the specification and simulation of processes by non programming experts. Nevertheless,

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 214–220, 2011.
© Springer-Verlag Berlin Heidelberg 2011

although the workflows can describe processes flow, they are not able to simulate interaction among different processes. Workflows are designed to be orchestrated. The orchestration assumes that the flow is managed by a central component that decides the next step in the process. In an interaction among processes the flow is decided in a distributed way. In this case, the next step is decided by the combination of the flows of the interacting components. That kind of execution of processes is usually known as process choreography [12]. Choreography assumes that the processes are able to exchange data to execute processes in a distributed way.

In this paper a scalable, customizable and update-able computer-aided sup-porting tool for a Human Interaction simulation environment is presented. This tool is based on choreography and orchestration principles. This solution is intended to be used by Interaction Designers and Usability Engineers at de-sign stages. This tool was developed in the framework of VAALID european project [1] and eMotiva Spanish government funded project [8].

2 Human Interaction Modeling Problems

The specification of a product design is usually approached by means of an exhaustive syntactical description of the states and the change rules that defines the device inherent behavior. This initial specification will allow product designers to share the product characteristics with the developers. Any error detected in this specification at later stages of the product lifecycle will have a direct impact in the form of an increase of development time and resources. This is because, the detected errors will force the redesign of the product and posteriorly return to start, again, the development process. In this way, the earlier the specification errors are detected, the lower impact to the final product costs they have.

The simulation of the initial specifications will allow to illustrate the behavior of the final product in defined contexts. This allows detecting design errors in early stages. The interaction modeling among simple devices has been approached in previous works through the use of Process Choreography techniques [6]. In this paper a system that is able to simulate the interaction of formally defined devices (using jPDL Workflow Language [3]) is presented. This framework al-lows detecting interaction problems that could affect the system execution at design time. This idea can be extended to human interaction modeling task in order to allow defining human behavior models that enable simulation systems to evaluate the accessibility of specific products by specific human groups. For example, this system will be able to validate a device designed to be used by people with special disabilities before the creation of the physical prototype. In this way, it will be possible to de ne general human interaction behavior models that represent different target groups and posteriorly use those models to validate products in design phases.

Nevertheless, the human behavior modeling is a more complex task than simulate the execution of devices. In most cases workflow languages have a very limited expressivity that can not describe the complex patterns that are usually found in human behavior [5]. To ensure that the system have enough capability to represent complex human behavior pattern it is needed to use high expressive and executable workflow languages. According to Workflow Patterns initiative [10], jPDL has a limited expressivity. Then, to enable the use of process choreography tool presented in [6] it was needed to improve the Workflow orchestrator in order to allow it to execute more expressive Workflows.

3 Timed Parallel Automaton for Human Interaction Simulation

In this paper, an improved Process Choreography tool using a very expressive Workflow Engine that allows professionals to simulate human interaction in presented. This tool will enable professionals to test the conceptual products in order to limit errors in design phase.

3.1 Timed Parallel Automaton for Human Behavior Modeling

To model human behavior relevant high expressive frameworks have been studied. One of the most expressive frameworks that have traditionally being used for process simulation are Petri Nets [7]. Although this framework is very expressive, their interpretation and execution is also complex. Other approaches based on automata theory are simpler. In addition, based on VAALID project experience, automata based design are more friendly for experts than Petri Nets approach.

For that reason an automata based Workflow Engine was developed for human behavior modeling. The selected approach was TPA (Timed Parallel Automaton) [4]. TPA satisfy the control flow workflow patterns ensuring a good expressivity. Also, TPA is able to de ne time based patterns that are widely used in human behavior models. The TPA formal defnition is presented as follows:

Definition 1. *A Timed Parallel Automaton(TPA) [4] is a tuple* $A = \{C, P, N, Q, T, \Phi, \Sigma, \gamma, q_0, F\}$ *where:*
C *is a finite set of clocks,*
P *is a finite set of actions,*
N *is a finite set of nodes where* $n \subseteq P \times C^{+} \forall n \in N$,
Q *is a finite set of states where* $q \subseteq N^{+} \forall q \in Q$,
T *is a finite set of time labels that can be generated by the Clock set* C,
Φ *is a finite set of actions results, called indicators,*
$\Sigma \subseteq T \cup \Phi^{+} \cup T \times \Phi^{+} \cup \{\lambda\}$ *is the finite input alphabet,*
$\gamma : N^{+} \times \Sigma^{+} \rightarrow N^{+}$ *is the node transition function,*
$\delta : Q \times \Sigma^{+} \rightarrow Q$ *is the state transition function,*
$q_0 \in Q$ *is the initial state,*
$F \subseteq N$ *is the set of final states.*

The node concept represents an atomic action in the problem. A node is composed by an action and a set of clocks which are used to model the time. The clocks are reset when the node is reached. A state is composed by a set of nodes. The state concept defines the set of nodes that are active in a given moment. Only one state can be active at any given time.

The double transition function allows the expression of complex workflow patterns. The first function represents complex transitions from multiple nodes to multiple nodes. On the other hand, the second function, reduces the complexity grouping the nodes in states. The formal definition of de TPA behavior is:

Definition 2. *let $q, q' \in Q$; $n, n' \in N^+$; $a^k \in \Sigma^*$; S is the set of symbols that can be generated by n where $S \in \Sigma$ and:*

$$\delta(q, a^k) = q' \wedge \gamma(n, a^k) = n'$$

Then a TPA derivation is defined as follows:

$$\delta(q, a^k\bar{x}) = \delta(q', \bar{x}) \wedge \gamma(n, a^k\bar{x}) = \gamma(n', \bar{x})$$

where:

$$n \subseteq qC \wedge n' \subseteq q' \wedge a^k \in S^*$$

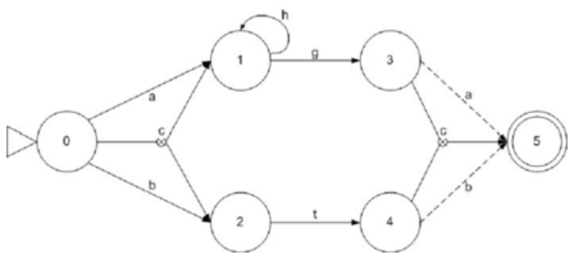


Fig. 1. TPA Example

Intuitively, the transition between states (q, q') occurs when a set of active nodes (n) generates a set of symbols (S) and there is a derivation when both functions and that connects the state q with the state q' and the set of nodes n with the set of nodes n' (respectively) with a set of symbols that are a subset of S. To finalize the transition, all of the active nodes in the set n are inactivated, and all of the inactive nodes of set n' are activated.

In a TPA, the Node transition concept represents the activation and inactivation of nodes in order to perform the execution of workflows. Moreover, the state transition concept is very important to the workflow representation because it offers a global view of the active nodes at any given time. Fig. 1 shows ex-ample of TPA. In the example, the arcs drawn with a continuous line are both state and node transitions. The arcs drawn with a discontinuous line are only node transitions. In the example, the a transition is available only when node 3 is active (state (3)). The b transition is available only when node 4 is active (state (4)). Finally, the c transition is available only when both nodes 3 and 4 are active (state (3,4)). This pattern is impossible to represent with only the information of the nodes. The state information helps to de ne TPA transitions of this kind. Therefore, state (3,4) indicates that the only transition available is the c transition.

The transition from node 1 to itself is also involved in the transition from state (1) to itself and state (1,2) to itself. In the case of state (1,2), if h transition res, only node 1 must be reactivated. If only state information is used, both nodes, 1 and 2 will be reactivated wrongly and therefore node transition information is also needed.

TPA are as expressive as safe Petri Nets keeping the complexity at Regular languages level [4]. This combines a high expressivity with a easily executable framework.

3.2 TPA Engine for TPA Execution in the VAALID Choreographer

In order to Execute TPA based automatons a module called TPAEngine was created. TPA Engine is a software tool that allows the execution of activities described as TPAs. TPA Engine is connected to VAALID choreographer and is able to make native calls to services registered in the choreographer. This enables the Orchestration system to simulate more complex processes to represent human behavior. In fig. 2 the architecture of TPAEngine is presented.

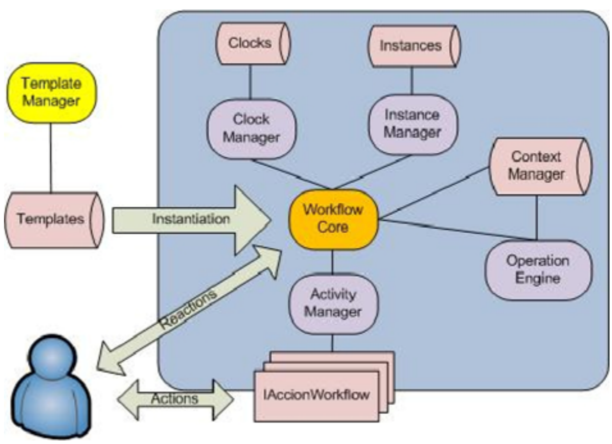


Fig. 2. TPA Engine Architecture

The TPA Engine is composed by the following modules:

- **Clock Manager:** The Clock Manager is the module that allows TPA Nodes to instantiate new clocks in the system. The Clock Manager has a clock repository where they are stored. When a clock elapses the clock manager notify the owner node and is deleted from the repository.
- **Instance Manager:** The Instance Manager is the on-execution time In-stance repository. This module stores and manages the workflow instances that are registered in the system while they are alive.
- **Activity Manager:** The activity Manager is in charge of processing the node assigned actions. This module is called by the nodes when they are reached. In that moment the activity manager acts as a broker with the choreographer. Using this system, the nodes can be assigned to services simulating user actions.

- **Context Manager:** The Context Manager is a data repository where the internal data of on execution instances are stored. This module allows workflows to perform memory based actions associated to previous actions' results
- **Operation Engine:** The operation Engine is the module in charge of compiling just-in-time the transition operations to be performed in the TPA. This module select the transitable derivations depending on the current status of the TPA instance and the transition condition based on boolean expressions.
- **Core Engine:** The core engine is in charge of the coordination of all the modules allowing the communication among the modules. It's also in charge of the instantiation of TPA templates, Node Clocks, resume the paused in-stances, select the transitions to be executed and coordinate the activity manager calls

The TPA Engine is connected to VAALID platform to allow the simulation of human behavior. In Fig. 3 is shown the VAALID application that can be used by design experts. Using this tool they can describe processes to simulate the interaction by using choreography technics among services, devices and human behavior models.

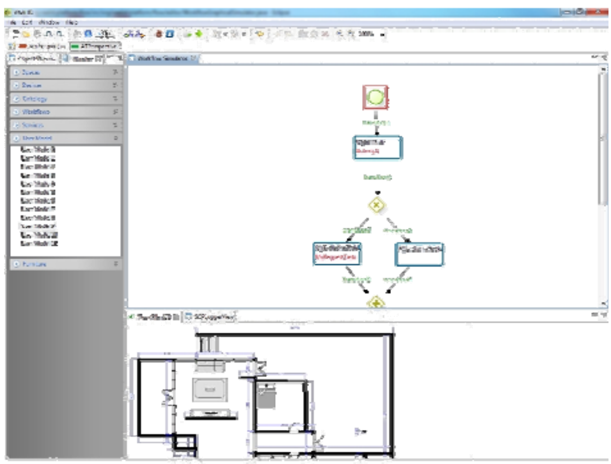


Fig. 3. VAALID Tool

4 Conclusions

In this paper, an enhanced framework for human behavior interaction simulation based on orchestration and choreography techniques is presented.

The use of those techniques has been tested in the simulation interaction among services and devices. The improvement proposed in this work allows the incorporation of very expressive process descriptions that can be used to incorporate human behavior models.

The use of human behavior models will allow design experts to simulate their products before the creation of prototypes. This makes them able to detect errors in early stages of the products lifecycle.

References

1. VII Framework Program IST Project 224309. Vaalid project:accessibility and us-ability validation framework for aal interaction design process (2008-2011)
2. COST consortium. Towards an inclusive future - Impact and wider potential of information and communication technologies. COST 219ter, Brussels (2007)
3. JBoss Red Hat Division. Jboss jbpm white paper, `http://jboss.com/pdf/jbpmwhitepaper.pdf`
4. Fernandez-Llatas, C., Benedi, J.M.: Activity-based work ow mining: A theoretical framework. In: Workshop on Technologies for Healthcare and Healthy Lifestyle (2006) ISBN:978-84-611-1311-8
5. Fernández-Latas, C., Lázaro, J.P., Benedí, J.M.: Workflow mining application to ambient intelligence behavior modeling. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5615, pp. 160–167. Springer, Heidelberg (2009)
6. Fernandez-Llatas, C., Mocholi, J.B., Sanchez, C., Sala, P., Naranjo, J.C.: Process choreography for interaction simulation in ambi-ent assisted living environments. In: Proceedings of XII Mediterranean Conference on Medical and Biological Engineering and
7. Murata, T.: Petri nets: Properties, analysis and applications. Proceedings of the IEEE 77(4), 541–580 (1989)
8. Turismo y Comercio Plan AVANZA I+D. Ministerio de Industria. Proyecto emo-tiva (2009-2011)
9. Story, M.F., Mueller, J.L., Mace, R.L.: THE UNIVERSAL DESIGN FILE Designing for People of All Ages and Abilities. The Center for Universal Design (1998)
10. van der Aalst, W.M.P., Barros, A.P., ter Hofstede, A.H.M., Kiepuszewski, B.: Workflow patterns. In: Distributed and Parallel Databases, p. 70 (2003)
11. WfMC. Workflow Management Coalition Terminology Glossary. WFMC-TC-1011, Document Status Issue 3.0 (1999)
12. Zimmermann, O., Doubrovski, V., Grundler, J., Hogg, K.: Service-oriented architecture and business process choreography in an order management scenario: rationale, concepts, lessons learned. In: OOPSLA 2005: Companion to the 20th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, pp. 301–312. ACM, New York (2005)

Building Multimodal Interfaces Out of Executable, Model-Based Interactors and Mappings

Sebastian Feuerstack and Ednaldo Pizzolato

Universidade Federal de São Carlos, Departamento de Computacão,
Rod. Washington Luis, km 235 - SP, Brazil
sfeu@cs.tu-berlin.de, ednaldo@dc.ufscar.br

Abstract. Future interaction will be embedded into smart environments offering the user to choose and to combine a heterogeneous set of interaction devices and modalities based on his preferences realizing an ubiquitous and multimodal access. We propose a model-based runtime environment (the MINT Framework) that describes multimodal interaction by interactors and multimodal mappings. The interactors are modeled by using state machines and describe user interface elements for various modalities. Mappings combine these interactors with interaction devices and support the definition of multimodal relations. We prove our implementation by modeling a multimodal navigation supporting pointing and hand gestures. We additionally present the flexibility of our approach that supports modeling of common interaction paradigms such as drag-and-drop as well.

1 Introduction

Future interaction will be embedded into smart environments offering the user to choose and to combine a heterogeneous set of interaction devices and modalities based on his preferences realizing an ubiquitous and multimodal access.

Such a flexible multimodal interaction requires user interface dialogues that are adaptive regarding the utilized modalities and their alteration. But current multimodal interfaces are often implemented for a predefined interaction hardware setup. Like for instance for the common mouse and keyboard setup of a computer, for a controller of a game console that supports gesture detection, or for the control wheel with speech feedback for navigating through the menus of a cockpit control in cars.

To tackle the complexity of designing and implementing multimodal interfaces, recent research has been focused on three main aspects. First, by model-driven user interface development (MDDUI) that describes a process for the tool-driven design of multi-platform interfaces though several models. MDDUI projects, such as [1] demonstrated basic multimodal applications. Second, executable models have been introduced into MDDUI [2, 3]. They enable to adapt the interaction modalities to the actual context of a user in a smart environment. Finally, the characteristics of different modalities and their relations have been investigated in depth [4] and platforms have been developed that support building multimodal interfaces out of components [5, 6]. To our knowledge these approaches of the first two categories support the design of

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 221–228, 2011.
© Springer-Verlag Berlin Heidelberg 2011

multimodal interfaces to a limited extent only. They enable modeling and adaptation between equivalent modalities and do not address multimodal fusion. The approaches in the third category are focusing on multimodal fusion between several modalities but restrict their findings to a particular use case for a fixed setup of modalities (like a cockpit for instance) or on command and control-interfaces.

In this paper we present our approach on modeling user interface dialogues with connected executable interactors. Each interactor consists of a set of declarative models that describe both, the structure as well as the behavior. Interactors are connected to form dialogues to support different multimodal hardware setups by multimodal mappings.

2 Related Work

Our work is based on the idea of combining earlier works about multimodal interaction and model-driven development of user interfaces to enable a developer to assemble multimodal interfaces based on pre-modeled interactors. Model-driven development has resulted in several connected design models that have been summed up by the Camelon Reference Framework [7] and in user interface languages such as USIXML [8]. But MDDUI has been applied to develop interfaces for pre-defined platforms only such as for example for XHTML+Voice browsers [1] or game consoles. Multimodal systems have been addressed by these approaches only to a very limited extend [1] and without addressing building interfaces out of complementary modes. Different to these approaches our interactor-based interfaces can be flexibly extended to new modes and media just by adding new interactors and mappings to a running system. Our approach is inspired by the findings of the iCARE platform [6] that supports building multimodal interaction out of components that are connected based on the CARE properties. These properties describe the relations between different modes, such as their complementary, redundant or equivalent combination.

Further on, we considered earlier work that we have implemented in the MASP platform [2] to support executing interfaces based on the designed models without the need for a further implementation phase. Different to the MASP that currently supports modeling equivalent multimodal relations only and is bound to the Eclipse Modeling Framework, we rely on state machines. State machines have been widely used in Case-Tools and are already standardized as part of UML and the W3C multimodal framework and therefore reduce the entry barrier for developers.

Finally, different to earlier work that applied state machines for the actual design of the interface dialogues, we use them to describe interactors. The interactor abstraction is mature [9] and has been for instance recently used to specify HTML and Java [10]. By using interactors, the interface design can be done in the same manner as it is actually done by UI builders to compose interfaces based on a toolkit.

3 Approach

Figure 1 illustrates the basic conceptual components of our approach as well as the relations between the models that we use to generate multimodal interfaces. Our approach supports the general MDDUI process that starts with a task analysis- and

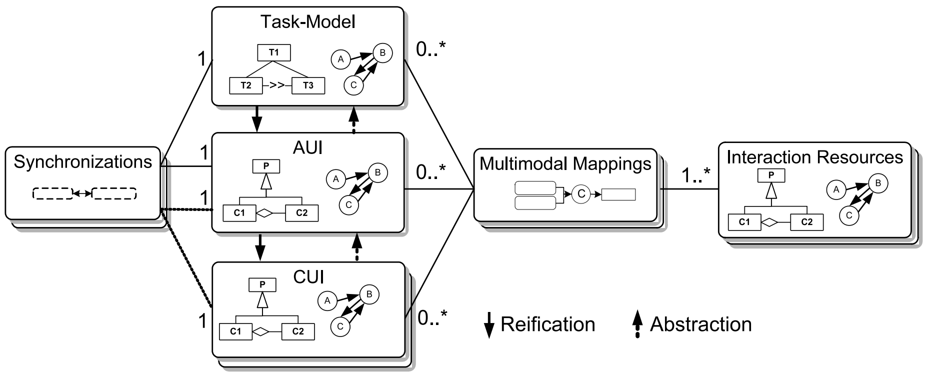


Fig. 1. The relations between the principal models

design to separate user from system tasks. An abstract user interface (AUI) model is used to specify the (modality-independent) interface elements that need to be realized for every interaction setup. The specific aspects of each concrete modality are captured by a concrete user interface (CUI) model. Additionally, and different to other approaches like CAMELEON[7] and USIXML[8], we describe the capabilities of the interaction resources (devices and modalities) in declarative models as well and target our approach to the design of multimodal interfaces that can be adapted to different modalities at system runtime. Further on we define mappings in a separate model and use them to glue all the models together at runtime [2].

Each design model is referring to interactors to describe an interface. Interactors are pre-defined components that describe the behavior of each model's atomic elements. We are using self-executable, reactive software agents to encapsulate interactors and specify their behavior by state machines. In the task model design for example, the developer assembles application and interaction task interactors that consist of an own lifecycle at runtime (initiated, executing, stopped, inactive, and done) like described earlier for instance in [11].

At design time, state machines (which are specified as declarative models) complement the user interface markup language with a behavior specification, which is currently missing in actual approaches like e.g. USIXML [8]. State machines are a well-known and a straight-forward way of describing and running reactive systems. They have been applied in several case tools that can generate source code from UML state charts[1] and will be standardized by the W3C to support speech-application design soon[2].

In the following sections we describe how multimodal applications can be designed and implemented with our Multimodal Interaction (MINT) Framework[3] and

[1] OMG Unified Modeling Language (OMG UML), Superstructure Version 2.2". http://www.omg.org/spec/UML/2.2/Superstructure/PDF/, last accessed 20/12/2010

[2] State Chart XML (SCXML): W3C Working Draft 16 December 2010 http://www.w3.org/TR/scxml/, last accessed 20/12/2010

[3] The MINT platform is available as open source at http://www.multi-access.de, last accessed 20/12/10

subsequently enhanced to support different multimodal setups. As a running example we focus on a multimodal dialogue navigation prototype. To proof our approach we applied several common interaction patterns to evaluate if they can be implemented with our approach. One of them, the drag-and-drop paradigm, will be presented in the end of this paper. The MINT framework does not require starting with a particular model. Thus, we start our example by modeling a concrete (graphical) interface and add the AUI layer for integrating further modalities thereafter.

3.1 Modeling Navigation for Graphical Interfaces

Modeling a graphical interface requires the composition of screens by selecting appropriate interactors like buttons, combo-boxes or menus for instance, which we call Concrete Interaction Objects (CIO). Our basic interactor for graphical concrete interface elements is depicted by the state machine in figure 2a.

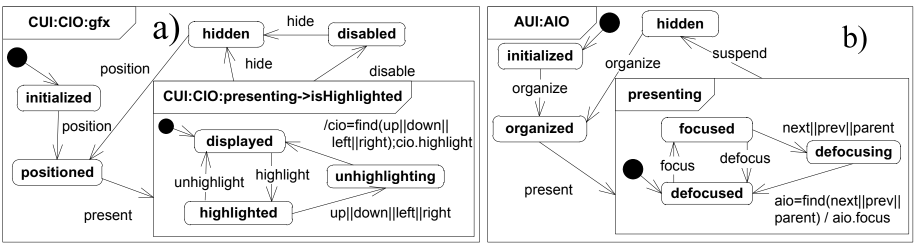


Fig. 2. a) Graphical CIO state machine. b) AIO state machine.

We support inheritance for interactor design and therefore the CIO interactor serves as the basis for all other, more specific interactors of a graphical interface. The basic life-cycle consists of a positioning, presentation, hidden and disabled phase. In the positioning phase, each CIO interactor calculates its absolute screen coordinates by identifying its neighbors that can be reached by interface navigation. During the presentation phase the interactor is part of the active user interface (UI) and reacts on navigation events. The user can therefore navigate to the interactor that gets then "highlighted".

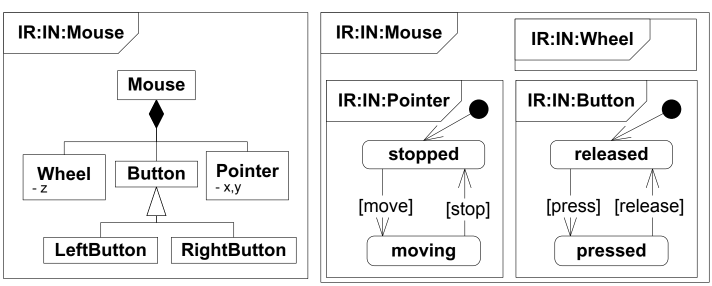


Fig. 3. The class diagram and state chart of a mouse

After the first CIO interactor has been highlighted, up, down, left, and right events can be sent to an interactor that is highlighted to navigate to the one. Each of those events "unhighlight" the receiving interactor, let it search the event's target interactor to that a "highlight" event will be sent.

For navigation several modalities can be considered as appropriate. The most basic one is a mouse that we design as a composed interaction resource (IR), which consists of a set of more basic interactors: a wheel, two buttons and a pointing interactor like shown in fig. 3. The behavior part of the mouse could be defined straight-forward by two state machines characterizing the pointer and a button. The pointer could be in the states "moving" or "stopped". While the pointer is in the "moving" state it communicates its x and y coordinates. The pointer is considered to be stopped if there is no movement for a particular time span. In the same manner, a mouse button can be described by a state machine to communicate its' two states: "pressed" and "released".

3.2 Mapping Specification for UI Element Highlighting while Pointing

Like depicted in fig. 1 we use mappings as the glue to combine the AUI, CUI and interaction resource specifications. The mappings rely on the features of the state machines that can receive and process events and have an observable state. Thus, each mapping can observe state changes and send events.

Fig. 4a shows the principal mapping we are using to change the highlighted graphical UI elements based on the actual pointing position. A mapping consists of Boxes with rounded and sharp edges. The former one define reactions on state changes of an interactor, the latter ones define system function calls or events. The mapping of figure 4a observes the state of the Pointer IR and gets triggered as soon as the pointer enters the state "stopped". The "C" specifies a complementary mapping, which requires all inputs of the C to be resolved for the mapping to start. The second mapping "assigns" (A) the observation of a moving pointer to an event that ensures that no CIO is highlighted during pointer movements.

Fig. 4. a) Basic pointer mappings. b) CUI with AUI synchronization

Therefore, as soon as the pointer has been stopped and coordinates of the stopped pointer could be retrieved, the findCIO function is called to check if there has been a CIO positioned at these coordinates and if it is currently not highlighted. The complementary mapping only executes if all three conditions can be evaluated and in our example fires a "highlight" event to the corresponding CIO.

3.3 Connecting Further Modalities

To add further, no-graphical media like sound or modes like gestures to control the interface we require the AUI model to keep them synchronized. Figure 2b depicts the basic abstract interactor object (AIO) that serves as a base for all other abstract interactors. Whereas it is quite similar to the basic CIO of figure 2a its semantics are different and independent of any modes or modalities. Thus, the AIO state machine processes the ordering of elements for a basic navigation. It supports browsing through all elements using "previous", "next" and "parent" commands (in contrast to the graphical-oriented navigation that supports directions like "up"/"down" based on their calculated coordinates). Further on, the highlighting feature of the CIO becomes a "focus" in the AIO describing that the actual interactor is in the focus of the user. Now, both interactors can be synchronized using the two bi-directional mappings of fig. 4b.

Now that we are having the navigation synchronized between the graphical CIO and the AIO we have two options to connect further modes and media. First, we can attach them to the AUI interactors so that we can benefit from the already existing synchronizations between AUI and CUI. Second, we could attach them to another CUI and add further synchronizations between the new CUI model and the AUI. The mappings shown by fig. 5a add gesture-based navigation and sound media and implement the former case and connect directly to the AUI. They require a gesture control interactor that supports two modes: When using two hands, like depicted in fig. 5b, the distance between the hands is measured and interpreted as "narrowing" or "widening". With only one hand detected, the interactor distinguishes two postures: one for issuing "next" and another one for issuing a "previous" command.

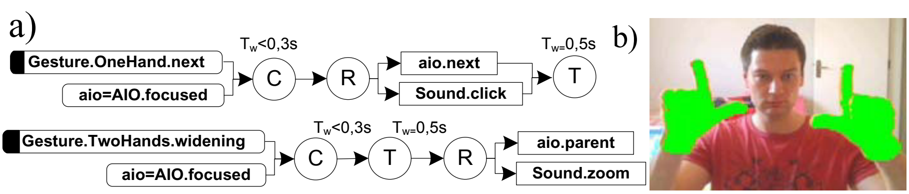


Fig. 5. a) Mappings to connect sound media and a gesture-driven control to the interface. b) Zoom gesture for widening and narrowing the user's focus.

The first mapping of fig. 5a waits for the next posture to appear while an interactor is in the state "focused". If this is the case the mapping gets executed and sends to events: A "next" event to the AIO (aio) that is in the focus of the user and a "click" event to a sound interactor that can generate a "click" sound. Both events are specified as redundant (R), which requires both by processed successfully. Alternatively they could be marked as equivalent (E), which requires only at least one of them (at least the aio.next event) to be processed successfully.

After both events have been fired the mapping waits (T) for half a second to re-initiate itself. Thus, if the user remains his hand in the "next" posture, the mapping gets fired every half second. The second mapping of fig. 5a to widen the focus to the parent interactor works in the same manner but does not require a timeout. A threshold inside the gesture interactor defines at which sensitivity the "widening" event is fired (this depends on the camera resolution and the users ability to keep the distance between their hands stable).

3.4 Mappings to Specify Interaction Paradigms

Our approach of describing interactions of composed interactors based on state machines and mappings turned out to be very flexible. Not only multimodal relations can be addressed by the mappings but also interaction paradigms like a "double-click" or "drag-and-drop" as well. Figure 6 depicts a mapping that specifies the drag-and-drop functionality for elements (AIChoiceElement) of an abstract list (AIChoice) on the AUI model level that is bound to the left button of a mouse. The AISingleChoice interactor (figure 7a) introduces a parallel super state to the presenting state of the AIO interactor. Additionally to the capability of gaining the user's focus, this super state describes the ability of a list to receive list elements that are dropped to the list.

Like shown in figure 7b, the dragging feature is not part of the list interactor but implemented by the interactor that describes an individual list element's behavior. There, we introduce a parallel super state as well that adds the capability of an element do be chosen and dragged. An element can only be chosen, if it is in the user's focus and it takes care, that all other list elements get "unchosen" if they are part of a single choice list. Given these interactor specifications, the drag-and-drop mapping of figure 6 is easy to understand: It waits for the left mouse button to be pressed and a list of AIChoiceElements that are in the state "chosen" to issue a "drag" event to them. As soon as the button is released it fires the drop event to the list that is currently in the user's focus.

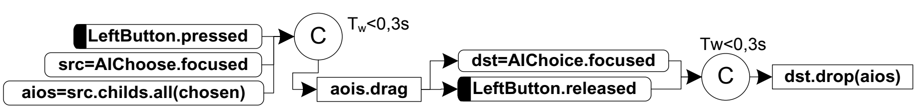


Fig. 6. Basic Drag-and-Drop mapping on the AUI model abstraction level

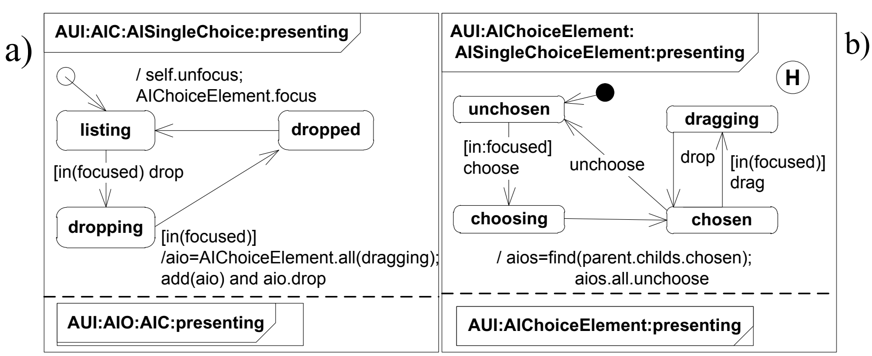


Fig. 7. a) AISingleChoice, the container for all SingleChoiceElements. b) SingleChoiceElement AIO state machine

Acknowledgments. The author is grateful to the Deutsche Forschungsgemeinschaft (DFG) for the financial support of his work.

References

1. Stanciulescu, A., Limbourg, Q., Vanderdonckt, J., Michotte, B., Montero, F.: A transformational approach for multimodal web user interfaces based on usixml. In: ICMI 2005: Proceedings of the 7th International Conference on Multimodal Interfaces, pp. 259–266. ACM Press, New York (2005)
2. Blumendorf, M., Lehmann, G., Feuerstack, S., Albayrak, Ş.: Executable models for human-computer interaction. In: Graham, T.C.N. (ed.) DSV-IS 2008. LNCS, vol. 5136, pp. 238–251. Springer, Heidelberg (2008)
3. Demeure, A., Calvary, G., Coninx, K.: COMET(s), A software architecture style and an interactors toolkit for plastic user interfaces. In: Graham, T.C.N., Palanque, P. (eds.) DSV-IS 2008. LNCS, vol. 5136, pp. 225–237. Springer, Heidelberg (2008)
4. Bernsen, N.O.: Multimodality theory. In: Tzovaras, D. (ed.) Multimodal User Interfaces, Signals and Communication Technology, pp. 5–29. Springer, Heidelberg (2008), doi:10.1007/978-3-540-78345-9_2
5. Lawson, J.-Y.L., Al-Akkad, A.-A., Vanderdonckt, J., Macq, B.: An open source workbench for prototyping multimodal interactions based on off-the-shelf heterogeneous components. In: EICS 2009: Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 245–254. ACM, New York (2009)
6. Bouchet, J., Nigay, L., Ganille, T.: The ICARE Component-Based Approach for Multimodal Input Interaction: Application to real-time military aircraft cockpits. In: Conference Proceedings of HCI International 2005, the 11th International Conference on Human-Computer Interaction, Las Vegas, Nevada, USA. Lawrence Erlbaum Associates, Mahwah (2005)
7. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A unifying reference framework for multi-target user interfaces. Interacting with Computers 15(3), 289–308 (2003)
8. Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., López-Jaquero, V.: USIXML: A language supporting multi-path development of user interfaces. In: Bastide, R., Palanque, P.A., Roth, J. (eds.) DSV-IS 2004 and EHCI 2004. LNCS, vol. 3425, pp. 200–220. Springer, Heidelberg (2005)
9. Duke, D., Faconti, G., Harrison, M., Paternó, F.: Unifying views of interactors. In: AVI 1994: Proceedings of the Workshop on Advanced Visual Interfaces, pp. 143–152. ACM, New York (1994) ISBN:0-89791-733-2
10. Trætteberg, H.: Dialog modelling with interactors and UML statecharts - A hybrid approach. In: Jorge, J.A., Jardim Nunes, N., Falcão e Cunha, J. (eds.) DSV-IS 2003. LNCS, vol. 2844, pp. 346–361. Springer, Heidelberg (2003)
11. Bomsdorf, B.: The webTaskModel approach to web process modelling. In: Winckler, M., Johnson, H. (eds.) TAMODIA 2007. LNCS, vol. 4849, pp. 240–253. Springer, Heidelberg (2007)

The First Interaction Design Pattern Library for Internet of Things User Created Applications

Marc Godon, Mohamed Ali Feki, Marc Roelands, and Lieven Trappeniers

Alcatel Lucent Bell Labs, Coperniculsaan 50, 2018 Antwerp, Belgium
{Marc.Godon,Mohamed_Ali.Feki,Marc.Roealands,
Lieven.Trappeniers}@Alcatel-Lucent.com

Abstract. In this paper, we report our analysis of extracting relevant existing and new interaction patterns that are candidates as enabling paradigms to facilitate Internet of Thing user created application building. We first define the context and underline what is an internet of thing user created application and what are the main research issues. We stress the focus on Interaction design as a must have paradigm to reach the Internet of thing user created application vision and highlight the research scope. In this paper we contribute with a template based interaction pattern that refers to competitive advantages and limitations with regard to our vision. The research method allowed us to sort out our first library of interaction pattern in this field. We conclude the paper with lab experimentation and lessons learned.

1 Introduction

The term "Internet of Things" (IoT) describes a number of technologies and research disciplines that enable the Internet to reach out into the real world of physical objects. Technologies like RFID, short-range wireless communications, real-time localization, and sensor networks are now becoming increasingly common, bringing the Internet of Things into commercial use. However, beyond the connectivity aspects of every thing or objects, less attention is paid to the value of what the Internet of Things could bring to the user experience. Indeed, we believe that the success of the internet of things will only happen when this phenomenon will actually bring significant value to users and society. Motivated by this vision, we started a new research activity, and proposed a “Do-it-Yourself” project (DIYSE[1]) in which end-users get the tools and the support to create and share their own smart applications and experiences. Our main challenge in this ongoing research track is to let end-users create their own applications for their self-defined smart environment. Therefore, they need to participate in the creation process, where they have the power and control over the creation of and use of the application. Making the parallel with Web2.0 and Web3.0 trends, we expected that the key to open the Internet of Things for mass creativity is depending on: open standards, easy application creation, rich interaction means, and End-User Programming. While a lot of works related to End-User Programming [1,2,3,4] is put in advance by recognized researchers, on which we can leverage, we are concerned that the user-thing interactions

[1] DIYSE project: Do-It-Yourself Smart Experiences http://www.dyse.org

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 229–237, 2011.
© Springer-Verlag Berlin Heidelberg 2011

need to be enhanced, and the user-environment interaction would need to be rethought. In fact, bringing the user to the application creation loop in an Internet of Things world will imply: considering existing interaction patterns and End-User programming methods and evaluate them from user control point of view; taking into account; wide segments of users should easily pick up well-defined abstract interaction patterns and make their stimulating experience smart.

In the rest of the paper, we introduce the interaction space in IOT environment from different perspectives. Then, we present our template and discuss the first five interaction patterns we found important for “User created application” research issue.

2 Interaction Space in IOT Environment

Interactions via creative, communicative, social, and functional interfaces, allows smart collaborations between smart objects, users, and their environment. From the objects perspective, users are facilitated to explore phenomena through introspection of smart object roles and collaborations. We agree with Gentry Underwood [8], that we are still far away from smart emphatic interactive objects and phenomena, which design for social, anthropological, and ecological arenas. Or as Gillian Crampton Smith states it as follows: “In terms of perceptual psychology, we’re starting to understand the functional limits of interaction between people and devices or systems: speed of response, say, or the communicative capacity of a small screen. But at the symbolic level of mood and meaning, of sociability and civility, we haven’t quite achieved the breathtaking innovativeness, the subtlety and intuitive “rightness,” of Eisenstein’s language of montage” [9].

We collaborate with open design communities and open DIY communities and research the need for new interaction paradigms empowering creativity and sociability at individual and community level as a starting direction in our search for this “rightness”. Many innovative projects are popping which explore the balance between the application of new technologies, functionality, need, self-expression, societal statement, empathy, beauty, and personal happiness (e.g. talking tree: a real-life immersive digital prototype on feelings and reflections of a tress somewhere in a street in Brussels, Belgium [2]). Fostering individual user and community creativity requested us to adopt a four-fazed creativity model: inspiration, divergence, exploration, and a convergence faze, where different interactions perspectives are elaborated. In this paper we limit our selves to the convergence phase of the creativity model but we are interested in opening the debate on the “whole” aspect in order to address the entire creativity model.

3 Template Based Interaction Pattern

3.1 Template

Inspired by Web design interaction pattern library namely Yahoo library[3] and Wellie[4], we believe that a interaction pattern library for Internet of Things user created

[2] http://talking-tree.com/
[3] http://developer.yahoo.com/ypatterns/
[4] http://www.welie.com/patterns/

application is an essential tool to enlarge the social network of users creating application and to reach the mass creativity goal we hope to achieve in our future research direction.

We tried to design a simple template that could tell what the interaction is about, what are the main features, what are the similar patterns. The template allows also giving examples, detailing the user creation flow and analyzing the pattern by extracting out the positive points as well as the limitations with regard to allowing non-skilled users to easily interact and create their desired application. The template elements are depicted in next figure.

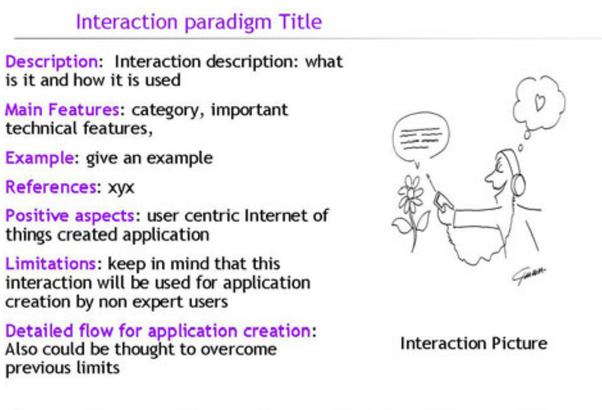


Fig. 1. Template interaction pattern for Internet of Thing user created applications

3.2 Research Methodology

In the first phase, we collected existing interaction paradigms from the Internet (e.g. forums, HCI communities, and books) and followed a series of brainstorm workshops for extra idea generation and selection. After that, we synthesized our finding in 10 interaction paradigms that we presented to a multidiscipline team composed of 10 researchers including computer scientists, creativity coach, designer, sociologist and business analyst. We shared the template and explained its main elements. We asked each researcher to look for new or known existing interaction paradigms and perform an interaction pattern analysis by using the template. After a collection and filtering phase (taking into account the multidiscipline team feedback as well as their individual contribution), we picked out a set of 5 interaction patterns that we will underline one by one in next section. Each analysis (next paragraph) should reflect very clear how non-skilled users would interact with the environment in order to create application using their Internet of Things daily objects. The five selected interaction patterns substantiate the kick-start of our interaction pattern library proposal.

3.3 Analysis: Interaction Patterns for Internet of Thing User Created Application

Interaction in IoT user created applications could be classified in different ways. We distinguish physical environment interactions from desktop-based interactions. Generally, in an environment interaction, end-users or developers instrument the environment and thus most of the interaction focuses on instrument objects to become smart (patterns 1 and 2). In a desktop-based interaction, end users are busy with filling into pre-programmed templates or wire visual widgets inputs and outputs (pattern 3 to 5). We noted that for IoT user created applications; a combination of the two interaction classes might be needed according to the application to be built.

Pattern 1: Smart Object Augmentation And Association. This interaction pattern is used to augment an object in order to empower or extend its capability (more processing, networking, situation aware capabilities, etc.). The augmentation is done with smart making enabling technology such as tags, labels, speakers, actuators, or sensing chips. For example augmenting a table with a speaker so the table could "speak". [5]

In the Smart-Its-project [5], the concept of friend relation ship between objects was introduced. Users could augment two objects and specify their relation ship by doing the same movement pattern (e.g. shaking them together or waving them simultaneously). The friend relationship can be used to create different kinds of applications. Making such type of objects associations is close related to physical object composition when users compose object to activate the object group behavior (see cubelets[5] and siftables[6]).

Example: By specifying the friend relationship, end users could elaborate a desired behavior of their personal objects. The behavior becomes only active when this friend relation ship is satisfied. For example: end users could connect their car key and credit card with a physical proximity relation so their credit card could not be used with the absence of their car key. (See Figure 2). Users could also use this interaction to create a child monitoring application by attaching two objects, associated with a friend relation ship, to both the parent and his kid.

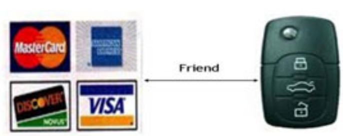


Fig. 2. Friend relation ship for smart object interaction

Detailed flow for application creation: Augment each object with a smart label and shake them simultaneously. As a result, an application creation template pops up on the screen of the user creation environment. The user has to specify the friend

[5] http://www.modrobotics.com/cubelets/
[6] http://alumni.media.mit.edu/~dmerrill/siftables.html

relationship between the objects (e.g. proximity) and the associated actions, application, and context to trigger the desired behavior of the objects.

Analysis: Object augmentation does not require system training and is easy and straightforward for creating simple applications. However, the programming paradigm is limited to consider events that could affect the object behavior in a specific context.

Pattern 2: Magnetic Poetry Interface. Description: This poetry inspired interactive application creation paradigm allows users to create applications in a way that takes advantage of the flexibility of natural language. It gives users the opportunity to build applications that reflect their way they linguistic envisage of the desired application, as opposed to requiring users to specify applications in terms of tangible devices.

Features: A graphical user interface for creating applications is augmented with magnetic cards, categorized in: what, where, when, and general, creates with its abstract vocabulary a poetic programming atmosphere.

Positive aspects: The user can express and combine concepts in a flexible but limited way. The restricted vocabulary allows easier system translation,

Limitations: Usually natural vocabulary based application requires huge translation efforts leading to solve conflicts (e.g. use of dictionaries).

Application Creation Flow: Magnetic poetry based descriptions are fed into the system by the end user, which, in turn, generates a specification that can be executed in a context aware environment (e.g. home environment).

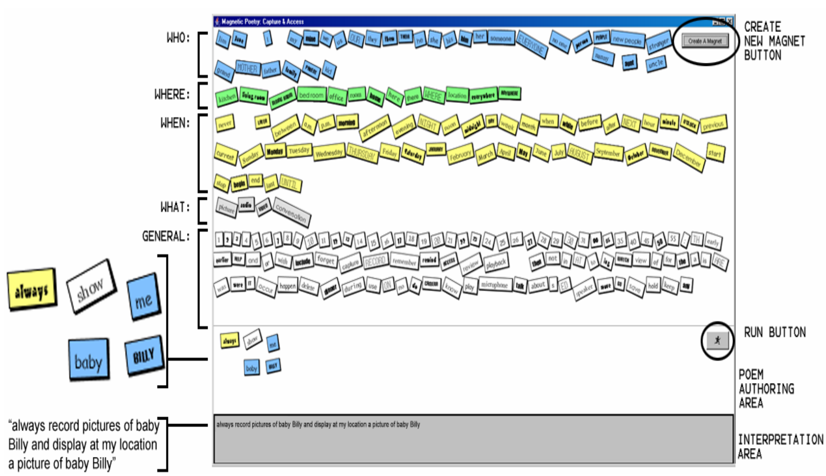


Fig. 3. Magnetic poetry [6]

Pattern 3: Pipeline Interaction. Description: Users can wire abstract concepts by linking their inputs and outputs of specific widgets in a graphical user interface. The widgets could represent multi-modal sensors and actuators, or components that are more complex.

Similar Paradigm: Workflow interaction,

Features: Low-level data transformation (filtering, threshold,) techniques and high-level data control mechanisms (synchronic function call, multicast, etc) are available at the same programming level (canvas).

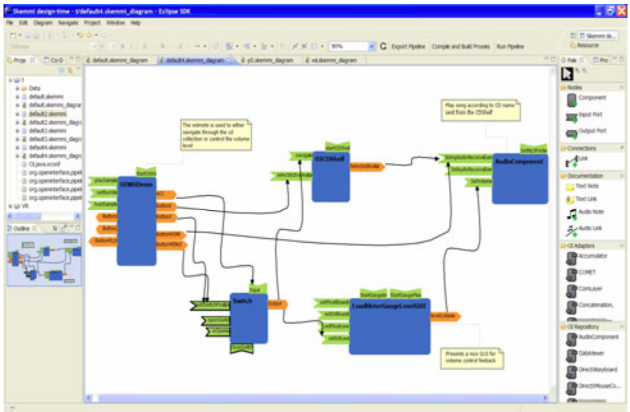


Fig. 4. Pipeline Interaction[7]

Positive aspects: Goal and task oriented creation paradigm is easy to comprehend by non-technical users

Limitations: The set of possibilities is limited and there is no support for collaborative design. This component composition technique requires technical knowledge of the components interface and behavior for fine-grained tuning of designed interactions. Illustrative examples are supportive used.

Application Creation Flow: Users can change iteratively concepts inputs and outputs. Documentation is attached to each component allowing easy understanding of components responsibilities, attributes, roles, and interfaces.

Pattern 4: Interaction by Example. Description: A user can visually inspect data, record desired behavior (e.g. complex gestures), uses it further as interaction pattern, and evaluate the effect runtime.

Features/Requirements: Signals inspectors e.g. used in oscilloscopes, computer vision techniques (such as OpenCV), and pattern matching algorithms (Dynamic Time Warping, five degrees-of-freedom gesture recognition) should be seamless integrated in an ready-to-use experience.

Example: The user performs an action, annotates its recorded signal, tests the generated behavior, and exports it to a supported tool.

[7] http://www.openinterface.org/platform/demonstrations

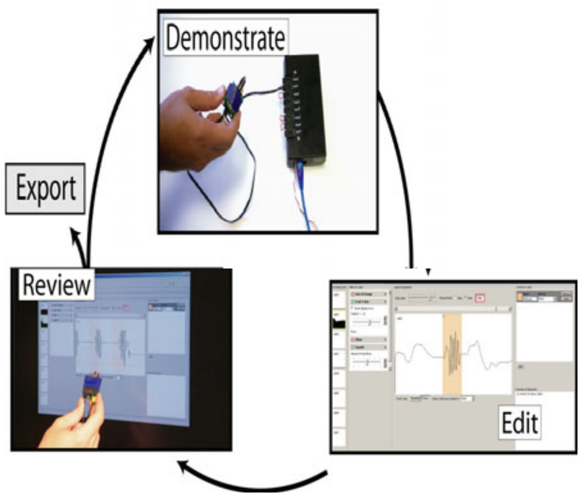


Fig. 5. Interaction by example[8]

Limitations: This application creation pattern is mainly applied for authoring sensor input. The demonstration is based on sensor data training and classification; where in some cases more than one training (demonstration) session is used.

Analysis: From a first analysis, there is a doubt how to apply the demonstration technique to recognize complex user activity requiring multiple data sources.

Positive points: seems perfect for single data source application, e. g pressing a force sensor.

Application Creation Flow: Users use a model to transform sensor inputs into application events, either discrete or continuous, by indicating thresholds or patterns on the incoming data. After that, they map events derived from sensor input on key presses or mouse events using the model, in order to operate a legacy application (e.g. map a wave gesture on the 'P' key to toggle playing/pausing in a media player application). Hence, the system will learn the demonstrated behavior for the demonstrator end user and will replicate the example when the same user will perform it.

Pattern 5: Tangible Metaphoric Story. Description: the interaction is constructed using tangible metaphors and stories.

Positive aspects: The Tangible metaphoric story allows end users to express their own meaning to interaction using tactile, visual, and metaphor embodied language. In addition, the tangible result permits a very fast funneling between detail and the whole with rich senses experienced. The metaphor brings to end-users an experience to support creativity [7].

---

[8] http://hci.stanford.edu/research/exemplar/

Limitations: It could be necessary to communicate cultural differences in advance or to take some learning steps in advance. Warm up sessions could be necessary to get used to the format in order to avoid resistance, hesitation to overcome.

Fig. 6. Tangible Metaphoric Story

4 Conclusion and Perspectives

The template we made allowed us to not only rethink the user environment interaction and enhance the user creation flow, but also to come up with a new interaction pattern that we experimented in our lab and we have a plan to make a functional prototype of them. The prototypes will be will be taken as input in the user research activities of the DIYSE project where expert and non-expert users will reflect on their effectiveness in their do-it-yourself social activities. The extended library will be presented in the conference.

From creativity point of view, the majority of patterns mentioned so far are best applied during the convergence phase of the creative process where ideas get concrete shape and designated interaction. We are currently working on patterns targeting the "fuzzy zone" of idea generation, the act of exploration for inspiration, and the desire for playfulness.

References

1. Sohn, T., Dey, A.: ICAP: an informal tool for interactive prototyping of context-aware applications. In: CHI 2003: CHI 2003 Extended Abstracts on Human Factors in Computing Systems, pp. 974–975. ACM, New York (2003)
2. Gross, T., Marquardt, N.: Collaborationbus: An editor for the easy configuration of complex ubiquitous computing environments. In: Proceedings of the Fifteenth Euromicro Conference on Parallel, Distributed, and Network - Based Processing - PDP, Bauhaus-University Weimar, Germany, IEEE Computer Society Press, Los Alamitos (2007)

3. Rodden, T., Crabtree, A., Hemmings, T., Koleva, B., Humble, J., Åkesson, K.P., Hansson, P.: Configuring the ubiquitous home. In: Darses, F., Dieng, R., Simone, C., Zacklad, M. (eds.) COOP, pp. 227–242. IOS, Amsterdam (2004)
4. Dey, A.K.: End-user programming: Empowering individuals to take control of their environments. In: ACM CHI, Workshop: The Future of User Interface Design Tools, HCI Institute, Carnegie Mellon University, ACM Workshop (2005)
5. Holmquist, L.E., et al.: Smart-its friends: A technique for users to easily establish connections between smart artefacts. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, p. 116. Springer, Heidelberg (2001) ISBN:3-540-42614-0
6. Truong, K.N., Huang, E.M., Abowd, G.D.: CAMP: A magnetic poetry interface for end-user programming of capture applications for the home. In: Davies, N., Mynatt, E.D., Siio, I. (eds.) UbiComp 2004. LNCS, vol. 3205, pp. 143–160. Springer, Heidelberg (2004)
7. Edith, A., Gauntlett, D., Weckstrom, C.: Defining Systematic Creativity. LEGO Learning Institute, Summary, PDF (2009), http://linkedith.kaywa.com/files/downloadAE7FA7078C209B8441A436A824E6E08A.pdf
8. Underwood, G.: Social Software : The Other " Design for Social Impact. Core77 Design Magazine and Resource, 1–11 (2009), http://www.core77.com/blog/featured_items/social_software_the_other_design_for_social_impact_by_gentry_underwood_15039.asp
9. Smith, G.C.: What Is Interaction Design? In: Moggridge, B. (ed.) Designing Interactions, pp. 8–19. The MIT Press, Cambridge (2007), http://www.designinginteractions.com/downloads/DesigningInteractions_foreword.pdf

Differentiating between Successful and Less Successful Products by Using MAInEEAC – A Model for Interaction Characterization

Steffen Hess, Andreas Maier, and Marcus Trapp

Fraunhofer Institute for Experimental Software Engineering, Fraunhofer-Platz 1
67663 Kaiserslautern, Germany
{steffen.hess,andreas.maier,marcus.trapp}@iese.fraunhofer.de

Abstract. Today, the success of a software product is defined by a great user experience caused by a well-defined interaction concept rather than features. Therefore we present a new interaction model called MAInEEAC (Model for Accurate Interaction Engineering, Enhancement, Alteration, and Characterization) that is able to show what makes an interaction a great experience and what are the elements of great human-computer-interaction. Having evaluated the ways of entering an address with nine distinct navigation systems, we present several findings which are crucial for making an interaction successful. Thereby it becomes clear that a higher level of detail is required in order to recognize important differences between single interactions. MAInEEAC provides these details and can be seen as a further step towards better understanding of human-computer-interaction.

Keywords: Human Computer Interaction, Human System Interaction, Interaction Design, Interaction Engineering, Interaction Model.

1 Introduction

Today, the way we interact with a software-intensive system is what sets apart successful products. A large number of features is no longer sufficient to define the success of modern software intensive systems. Users expect good usability and want to enjoy a great experience when interacting with the system. What makes, however an interaction a great experience? What are the elements of great Human Computer Interaction (HCI)? Enabling natural or intuitive interactions via (multi-) touch, natural language, free-form gestures, or multimodal interactions combining advantages of these rather new kinds of interaction might be a promising approach, but the number of interaction concepts increases more and more. Nevertheless, the actual elements of great HCIs and their dependencies are not sufficiently known yet. At the moment, several interaction models exist in the large field of HCI [1, 3, 4, 6, 7, 8, 9, 10, 11], which usually clearly have their origin in a specific HCI discipline such as computer science, anthropology, psychology, etc. When analyzing these models, we discovered that each of them has deficits in some points, mainly due to the emphasis of single aspects. A visual designer, for example, might use a model representing elements of

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 238–247, 2011.
© Springer-Verlag Berlin Heidelberg 2011

the graphical user interface as part of the system, whereas an interaction designer might use a model representing the influences of devices, media, and modalities on each other. Thus, a seamless workflow and well-grounded communication between these roles is hard to achieve when using different models. Nevertheless, aspects taken into account by each role involved in interaction engineering are crucial for developing a successful product. In order to engineer a new system that enables a great interaction experience, we need to know the elements that constitute such an interaction and their dependencies. Therefore, we need an interaction model that allows us to analyze the interaction concepts of successful systems.

Hence, we created a new conceptual interaction-centered HCI model. Our model is called MAInEEAC[1] (Model for Accurate Interaction Engineering, Enhancement, Alteration, and Characterization). Currently, we are using MAInEEAC as a descriptive model only. It allows us to characterize HCI precisely. This characterization may serve as a basis for HCI enhancement and alteration, allowing us to use MAInEEAC as a prescriptive model as well, which supports all roles involved in the interaction engineering process and therefore comprises all HCI details relevant for developing a successful product. In this paper, we briefly describe how to characterize an information system with the help of MAInEEAC. Furthermore, we characterize several different navigation systems in order to show how our model can help to find the aspects that let a product become successful and the shortcomings of less successful products. It becomes clear where the differences between those systems are in detail and why users prefer one system to another. Having analyzed the HCI concepts of successful systems, we will be able to engineer a new system that enables a great interaction experience by transferring these concepts to the new system. An evaluation of navigation systems by expert users showed some interesting findings. For example, we can show that there is a correlation between the number of interaction steps needed to accomplish an interaction, and the kind of interaction like gesturing or speaking. Furthermore, the detailed type of interacting with the system like tapping or clicking has a strong impact on the user experience. In addition, we show that the feedback on the user's input as well as the subjectively and objectively estimated time needed for performing the total interaction are crucial.

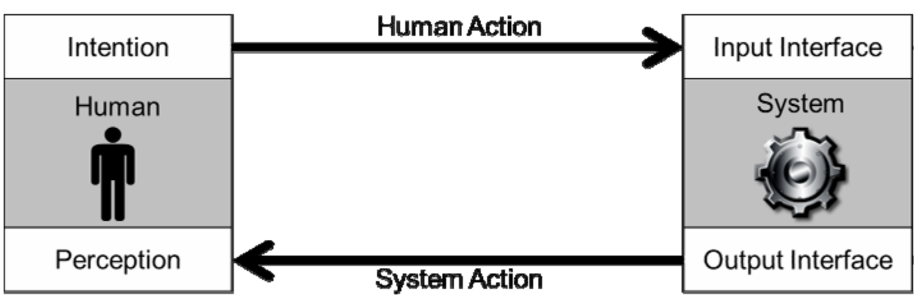


Fig. 1. MAInEEAC Overview

[1] MAInEEAC is supposed to be pronounced "maniac".

2 MAInEEAC in Detail

In general, MAInEEAC is built up view-based, focusing elements needed for HCI by a human, elements needed by a system, and elements building the interaction by connecting elements of human and system (cf. Fig. 1). In the scope of this paper are the *Interaction View*, the *Human Action View*, the *System View,* and the *System Action View*. Each view shows highly detailed aspects of HCI. Since MAInEEAC has been built from an interaction-centered point of view and since we are mainly interested in the details of information transmission between human and computer, this paper mainly focuses on the *Interaction View*, which comprises all elements forming an interaction (cf. Fig. 2). In addition, the *System Action View* (cf. Fig. 2 – lower part) emphasizes important HCI parts of the system and provides the elements that can be influenced by designers. A basic element of MAInEEAC is the *Elementary Interaction* which is strictly terminated. It is fulfilled when exactly one *Human Action* leads to an *Application Feedback* that meets the human's *Intention*. A *Human Action* comprises the whole activity of a *Human* that is performed in order to transmit information to a *System*. That activity is necessary to fulfill a human's *Intention* and is initiated by that *Intention*. Since a *Human Action* is influenced by the current *Environmental Context*, this context is described as an attribute of *Human Action*. An *Intention* causes a single *Human Action*, not a whole *Interaction* in the first instance.

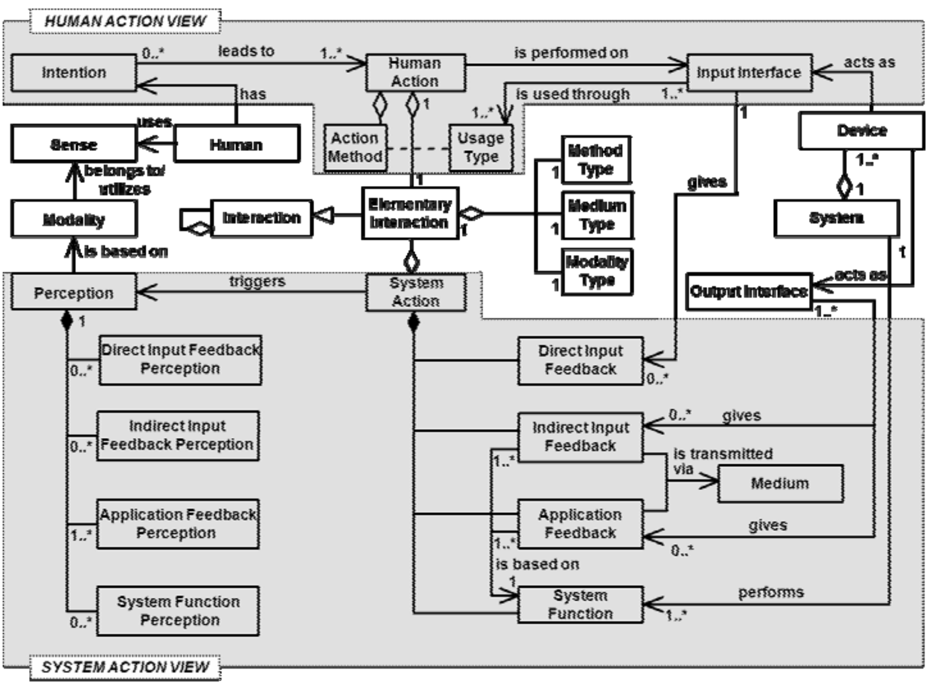


Fig. 2. Interaction View of MAInEEAC

That means that MAInEEAC supposes even the beginning of an *Interaction* on a more detailed level than most other HCI models. In the end, the whole *Interaction*, which might consist of other *Interactions*, is built up by single *Human Actions* and *Elementary Interactions* and might change dynamically during that process. Due to the definition of an *Elementary Interaction* as a fulfilled step towards the accomplishment of an *Intention*, MAInEEAC comprises the evaluation of actions implicitly. Other models like Norman's Seven Stages of Action [7] evaluate actions explicitly and do not offer strict evaluation rules.

2.1 Human Action View

The *Human Action View* (cf. Fig. 2 – upper part) describes the way the *Human* accomplishes his *Intention* on the *Input Interface* of the *System* (cf. top arrow of Fig. 1).

In MAInEEAC, this accomplishment is called *Human Action* and involves the whole activity performed by a *Human* to transmit information to a *System*. Every activity might be influenced by the *Environmental Context* in which the activity takes place. A *Human*, for example might not want to transmit information to the *System* via speech when he is in a noisy environment, or he might want to interact with the *System* from a distance when his environment is a huge warehouse. The *Environmental Context* is an attribute of *Human Action* and thus has to be specified when using MAInEEAC. Each *Human Action* consists of at least one *Action Method* which specifies what the action of the *Human* is. It comprises generic *Human* movements according to original *Human* abilities. We distinguish four classes of *Action Methods*: “fine motor skills”, “gross motor skills”, “facial expressions”, and “vocal utterances”. Each of these classes comprises particular and precise *Action Methods*. Furthermore, a *Human Action* always has a *Method Type* which specifies if the action is unimethodical or multimethodical. If there is only one kind of *Action Methods*, the type is unimethodical. If there are at least two kinds of action methods, the method type is multimethodical.

When performing the *Human Action*, every *Input Interface* features different *Usage Types* that specify exactly how an *Input Interface* is used to transmit information to the *System*. “Pressing object”, for example as an *Action Method* of the class “fine motor skills” is underspecified. There is a set of concrete movements being performed within a *Human Action*. It makes a big difference whether one presses the left or the right mouse button or if someone performs a single or a double click. In addition, the *Usage Type* concretely specifies the *Action Method*. If the *Usage Type* of the *Input Interface* is for example “single left click”, the *Action Method* is determined as “pressing object”. The *Usage Type* is the main aspect of a *Human Action* from a designer's point of view. No matter if the *Interaction* itself is designed or the user interface, the *Usage Type* is what has to be designed and what influences the design of other parts of the *Interaction*. By concentrating on that special part of a *Human Action*, the designer is able to better account for a successful *Interaction*.

2.2 System View

The *System View* describes the system side during HCI (cf. right side of Fig. 1). MAInEEAC does not treat the *System* as a black box like many other HCI models do

[7, 10]. The *System* components highly influence HCI and thus are shown in the same detail as *Human* elements and elements of the *Interaction* itself. The *System* communicates with the environment through its *Input* and *Output Interface*.

Every *Input Interface* is characterized by the *Usage Types* it offers. Furthermore, an *Input Interface* may give *Direct Input Feedback* which comes straight from the *Device* without the use of any *Medium* (for example, the sound that occurs when pressing a button on a mouse or keyboard). When a *Device* acts as an *Output Interface*, it gives *Application Feedback* and might give *Indirect Input Feedback* in addition. Both are transmitted via a *Medium*: *Indirect Input Feedback* is a reaction to the usage of the *Input Interface* to confirm the system's correct understanding of that *Human Action*, for example, highlighting of a selected menu item transmitted via the *Medium* "graphic". It is always based on a *System Function*. *Application Feedback* is the intended reaction of the *System* to the human action and confirms or declines the human *Intention*; for example giving the sound of paper being crumpled transmitted via the medium sound/auditory icon when the recycle bin is being emptied. *Application Feedback* is always based on a *System Function*. In contrast to the types of feedback, a *System Function* does not directly address the *Human* but can be perceived by the *Human* (for example, deleting files on a hard disk). Both *Indirect Input Feedback* and *Application Feedback* can be influenced by a designer. *Direct Input Feedback* and *System Functions* may only be influenced within strong limitations.

The benefit of this distinction is a clear differentiation of *System Actions* referring to a *Human Action* (*Indirect*, *Direct Input Feedback*) from those referring to the actual (intended) "reaction" of the *System* (*Application Feedback*, *System Function*).

When characterizing an *Interaction* using MAInEAAC, the elements *Action Method*, *Usage Type*, and *Input Interface* influence each other. That is, by making the dependencies of these elements explicit, the final interaction concept becomes consistent and interaction engineering is facilitated. When for example an *Action Method* is determined, the possible *Input Interfaces* and *Usage Types* are deduced from that *Action Method*. If, for instance, the *Action Method* "Natural Speech" is determined, the number of *Input Interfaces* possible for sound input is low and the possible *Usage Types* are restricted to that *Action Method*. The *Action Method* in conjunction with the *Usage Type* restricts the number of possible *Input Interfaces* even more. With respect to the *Usage Type*, the same holds for the *Action Method* in conjunction with the *Input Interface*. The *Usage Type* determines the *Action Method* and restricts the number of available *Input Interfaces* to those that offer the chosen *Usage Type* and allow for performing the deduced *Action Method*. When for example the *Usage Type* "Natural Speech" is selected, "Speaking" is automatically determined as the *Action Method*. The list of *Input Interfaces* is restricted to different microphones like desktop microphone, handheld microphone, wearable microphone, etc. When we decide on a particular *Input Interface*, the *Action Method* as well as the *Usage Type* are restricted at the same time. When *Input Interface* and *Usage Type* are determined, the *Action Method* is deduced from those. This does not hold for *Input Interface* and *Action Method*: When the *Input Interface* and the *Action Method* are determined, we still have a choice regarding the *Usage Type*. When for example we decide on "Microphone" as *Input Interface* and "Natural Speech" as *Usage Type*, "Speaking" is automatically determined as the *Action Method*. When we leave the

Usage Type open and decide on an *Action Method* from the class “Vocal Utterance” (e.g. “Speaking”), we can still decide on which *Usage Type* to apply (e.g. “Natural Speech” or “Command-based Speech”).

2.3 System Action View

The *System Action View* (cf. Fig. 2 – lower part) describes the way the *System* reacts to a *Human Action* and acts in order to transmit information to the *Human*.

The System Action is a composition of Direct Input Feedback, Indirect Input Feedback, Application Feedback, and System Function. Depending on the Environmental Context in which the System Action takes place, the System Action might be influenced by that context. For example, the System Action might not be to transmit information to the Human via sound, when it is in a noisy environment or to show relevant information to a particular person when it detects that that person is near.

The *System Action* triggers a *Human’s Perception* which is composed according to the *System Action*. Because of this segmentation, we can trace which *Perception* is triggered by which kind of *System Action*. For each kind of *Perception*, a *Modality Type* is specified that determines if the *Perception* is unimodal or multimodal. A unimodal *Perception* is given when exactly one *Modality* is used to perceive the part of the *System Action*. In a multimodal *Perception*, at least two different *Modalities* are used. In MAInEEAC, *Modality* refers to the *Modality* of *Senses*. Thus, we can adhere for the human's cognitive load. In addition to that, for *Indirect Input Feedback Perception* and *Application Feedback Perception*, a *Medium Type* is specified that determines if the *Perception* is unimedial or multimedial. This is necessary for those two types because they are the only ones that are transmitted via a *Medium* and thus influenceable by a designer. If one *Medium* is used for transmission, the *Medium Type* is unimedial; if at least two different *Media* are used, the *Medium Type* is multimedial. We refer to the definition of "medium" in ISO14915-3 [5]. In MAInEEAC, *Media* are the system's options of expression towards the *Human*. In order to reduce the *Human's* cognitive load and therefore to increase the products success, *Media* can be matched to the *Action Method* by a designer.

An *Interaction* vocabulary connected to MAInEEAC is used for the subjective evaluation of several dimensions of *Usage Type*, *Application Feedback*, and *Elementary Interaction* [2]. For the evaluation we present in this paper, we only use the dimension “speed” of that vocabulary.

This level of detail MAInEEAC provides is needed for showing the elements that differentiate successful products from less successful ones. On a higher level, on which most other models are settled, important coherences and dependencies cannot be recognized from our point of view.

3 Evaluation of Distinct Navigation Systems by Using MAInEEAC

In order to cover all details of MAInEEAC, we developed a Microsoft Excel template. By filling in that template, all relevant information for describing and evaluating the

Interaction with an information system is available and systems become comparable to each other. In detail, the template consists of the following parts: at first, general information about the context of use (human, intention, system, and devices of which the system consists) is given. The second part describes HCI in all details, namely the *Elementary Interactions*, the corresponding *Human* and *System Actions* with their subjectively measured dimensions, the *Action Methods* and their *Types*, the *Usage Types*, the *Input* and *Output Interfaces*, the *Direct*, *Indirect* and *Application Feedbacks* and their *Perceptions*, the *System Functions*, and the *Media* and *Modalities* used for giving feedbacks and perceived by the *Human*. In the third part, the aggregated *Elementary Interaction* and *Interaction Media*, *Modality*, and *Method Types* are given as well as the subjectively measured dimensions of the *Interaction* as a whole. At last, the duration of time needed for accomplishing the whole *Interaction* by an expert user is given. Except for the subjective data, the template has been filled out by ourselves before the evaluation has started.

During the evaluation, we had a look at nine distinct navigation systems, eight mobile ones and one in-car system. Table 1 shows the navigation systems we have evaluated, their types (mobile or in-car), and the devices and operation systems on which they run.

Table 1. Navigation Systems that have been evaluated

Vendor	System	Type	Device, OS
Google	Mobile Navigation	Mobile	HTC Desire, Android 2.2
Navfree	Navfree 1.6.0	Mobile	iPhone 4, iOS 4.2
Navigon	Mobile Navigator 1.2	Mobile	iPhone 4, iOS 4.2
Opel	DVD 800	In-car	
Roadee	Roadee	Mobile	iPhone 4, iOS 4.2
Skobbler	Skobbler 3.0.1	Mobile	iPhone 4, iOS 4.2
TomTom	Go 910 6.525	Mobile	TomTom Go 910 v1.21, Boot 4.86
TomTom	TomTom 4 iPhone	Mobile	iPhone 4, iOS 4.2
Vodafone	Find & Go	Mobile	iPhone 4, iOS 4.2

The evaluation of the systems was done by 10 expert users each (8 male, 2 female, average age 30,3, min 23, max 35 years old) who have been asked to only evaluate the handling of the address input, but not the navigation itself. Expert users for the address input could be found very quickly, because it requires only very short training of the address input in order to become an expert user. Before data have been gathered, each expert user had the possibility to get familiar with the current system that should be evaluated. After having evaluated each navigation system, the experts answered questions regarding the estimated value of the product and its estimated usage on a daily routine.

In terms of pure *Interaction* - and this is what we focus on in this paper - the main factor for differentiating successful products from less successful ones is the number of *Elementary Interactions* needed to fulfill an intention. The more *Elementary Interactions* are needed, the longer it seems to take to fulfill an *Intention*, both subjectively and objectively. In order to determine an *Elementary Interaction* and therefore influence the estimation of time needed for accomplishing an *Interaction*, *Application Feedback* is defined as the *System Action* expected by the user and needed to fulfill his *Intention*. *Application Feedback* is influenced by designers and thus the kind of *Application Feedback* representation (in this case one form for each information vs. a single screen requesting the information) as the second success factor influences the subjectively and objectively measured time for an *Interaction*. MAInEEAC allows designers to concentrate on *Application Feedback*, without being distracted by other (non-influenceable) elements of HCI, mainly by concerning the whole *System Action*. From the evaluated navigation systems, Vodafone Find & Go, Roadee, and Skobbler use a form requesting all information, the other navigation systems lead their users through single screens for each information. Furthermore, it is crucial to let the user know when *Application Feedback* is given in order to avoid that they become impatient or uncertain. The subjective measurement of time is the third factor of success and done by estimating the speed of *Action Methods*, the corresponding *Application Feedbacks*, and the *Interaction* as a whole. Objectively, the absolute time needed by an expert user for accomplishing the whole interaction is measured and builds the fourth success factor. Usually, a large number of *Elementary Interactions* seems to correlate with time both subjectively and objectively estimated as slow (cf. Table 2).

Table 2. Elementary Interactions and Time Measurements with each Navigation System

System	Number of elementary interactions	Average subjective time (slow=1/ fast=2)	Average objective time (in seconds)	Estimated value	Estimated usage on a daily routine
Google Mobile Navigation	3	fast (2)	16,94	25-50€	yes
Navfree 1.6.0	7	average (1,65)	17,45	<25€	no
Navigon Mobile Navigator 1.2	5	fast (1,95)	18,32	25-50€	yes
Opel DVD 800	14	slow (1,2)	47,88	>50€	yes
Roadee	4	slow (1,05)	45,64	<25€	no
Skobbler 3.0.1	4	average (1,5)	28,71	<25€	yes
TomTom Go 910 6.525	7	average (1,6)	39,22	25-50€	yes
TomTom 4 iPhone	6	fast (2)	24,36	<25€	yes
Vodafone Find & Go	4	average (1,65)	30,54	<25€	yes

But a low number of elementary interactions does not necessarily lead to time being estimated as “fast”. In contrast, a very high number of elementary interactions lead to a subjective and objective “slow” total interaction. Anyway, the subjective estimation of time may differ from the objective estimation. E.g. “average” subjective estimations of an *Interaction*, but fast objective ones (cf. Tomtom Go 910 with Navfree). We suggest that these characteristics are due to the kind of *Input Feedback* implemented by the interaction designer. The *Input Feedback* (*Indirect Input Feedback* as well as *Direct Input Feedback*) is the fifth factor of successful *Interaction*. Reasonably used, it leads the user's attention, increases his certainty and influences his subjective estimation of time. By splitting *Input Feedback* into *Direct* and *Indirect Input Feedback*, MAInEEAC not only shows all aspects of *Input Feedback*, but also differs *Input Feedback* being influenced by a designer (*Indirect Input Feedback*) from the one that cannot (*Direct Input Feedback*). Similar to the *Application Feedback*, the designer is able to concentrate on the *Input Feedback* separately with MAInEEAC. Another factor correlating with the number of *Elementary Interactions* and the time needed for accomplishing an *Interaction* is the *Action Method*: reasonably applied, this sixth factor of product success can minimize the number of *Elementary Interactions* and the total time needed immensely. From the navigation systems we have evaluated, Google Mobile Navigation is the only one providing speech input and thus can be used with the *Action Method* "Speaking". The seventh success factor is the *Usage Type*: we found a huge impact on both, the subjectively and objectively needed total time for accomplishing an *Interaction* depending on the *Usage Type* applied. The more direct the manipulation of data is (tap vs. knob rotation) the less time seems to be needed for the *Interaction*. From the navigation systems we have evaluated, the Opel DVD 800 navigation system is the only one that is controlled with a rotary knob, all others are controlled via touch screens or via voice. Since the directness of manipulation depends on the *Input Interfaces* (e.g. touch-screen vs. PC-mouse) and adequate *Output Interfaces* (e.g. 22" screen displaying 16.7 mio. colors vs. 2.4" monochrome display) in use, the user interface is the eighth factor of a successful product. The splitting of the whole *Human Action* into *Action Method*, *Usage Type*, and *Input Interface* is an important contribution of MAInEEAC to HCI design and comprehension from our point of view. Designers and evaluators can concentrate on one of these aspects and are supported with design, engineering, and characterization of the other aspects by that splitting when they use MAInEEAC.

4 Conclusion and Future Work

Currently, we are able to characterize the *Interaction* between a *Human* and a *System* precisely with MAInEEAC. These characterizations are complex at the time. We currently investigate facilitations that make the model more applicable. But we are already able to show important factors of success, due to MAInEEAC's high level of detail. The success factors we found with the evaluation described in this paper are shown in Table 3.

Table 3. Success factors found with MAInEEAC

No.	Success Factor
1	The number of *Elementary Interactions* needed for accomplishing an *Interaction*
2	The kind of *Application Feedback* representation
3	The subjective measurement of time needed for *Action Methods*, corresponding *Application Feedbacks*, and the whole *Interaction*
4	The absolute time needed for accomplishing an *Interaction*
5	The implementation of *Input Feedback*
6	The *Action Method(s)* provided
7	The *Usage Type* provided
8	The user Interface

Future steps will be to do more evaluations with MAInEEAC in order to approve the current results and to get statistically more significant data. All in all, we believe that MAInEEAC is a sophisticated model which takes us a step further on our way towards a better understanding of human-computer-interaction.

References

1. Bonto-Kane, M.V., St. Amant, R.: Computational modeling approaches help guide early design efforts for usability. In: The Fifth Richard Tapia Celebration of Diversity in Computing Conference: Intellect, Initiatives, Insight, and Innovations, TAPIA 2009, Portland, Oregon, April 01-04. ACM, New York (2009), doi: http://doi.acm.org/10.1145/1565799.1565802
2. Diefenbach, S., Hassenzahl, M., Kloeckner, K., Nass, C., Maier, A.: Ein Interaktionsvokabular: Dimensionen zur Beschreibung der Ästhetik von Interaktion. In: Brau, H., Diefenbach, S., Göring, K., Peissner, M., Petrovic, K. (eds.) Usability Professionals 2010, pp. 27–32. German Chapter der Usability Professionals' Association e.V, Stuttgart (2010)
3. Dubberly, H., Pangaro, P., Haque, U.: ON MODELING What is interaction?: are there different types? Interactions 16(1), 69–75 (2009), doi: http://doi.acm.org/10.1145/1456202.1456220
4. Goldbogen, G.: A generalization of the GOURD model of human-computer interaction. In: Proceedings of the 37th Annual Southeast Regional Conference (CD-Rom) ACM-SE 37, p. 48. ACM, New York (1999), doi: http://doi.acm.org/10.1145/306363.306424
5. ISO 14915-3. Software ergonomics for multimedia user interfaces– media selection and combination. Draft International Standard (2002)
6. Kolski, C.: A "call for answers" around the proposition of an HCI-enriched model. SIGSOFT Softw. Eng. Notes 23(3), 93–96 (1998), doi: http://doi.acm.org/10.1145/279437.279483
7. Norman, D.A.: The Design of Everyday Things. Basic Books, New York (1988)
8. Obrenovic, Z., Abascal, J., Starcevic, D.: Universal accessibility as a multimodal design issue. Commun. ACM 50(5), 83–88 (2007), doi: http://doi.acm.org/10.1145/.1241668
9. Pastel, R., Skalsky, N.: Object-action association: a HCI design model. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, IUI 2004, Funchal, Madeira, Portugal, January 13-16, pp. 295–297. ACM, New York (2004), doi: http://doi.acm.org/10.1145/964442.964510
10. Schomaker, L., Munch, S., Hartung, K.: A taxonomy of multimodal interaction in the human information processing system. Technical report, ESPRIT BRA, No. 8579 (1995)
11. Sutcliffe, et al.: A Method and Advisor Tool for Multimedia User Interface Design. International 1230819 Journal of Man-Machine Studies 64(4), 375–392 (2006)

Patterns for *Usable* Accessible Design

Homa Javahery[1], Michael Gower[1], Daniel Sinnig[2], and Peter Forbrig[2]

[1] IBM Pacific Development Centre, Burnaby, B.C., Canada
[2] Computer Science Institute, University of Rostock, Germany
{homa,gowerm}@ca.ibm.com,
{dasin,pforbrig}@informatik.uni-rostock.de

Abstract. Accessibility barriers exist on most websites and systems, making them difficult to use for individuals with disabilities and other functional limitations. *Usable* Accessible Design aims to design usable software that addresses the potential needs of all users in the user-centered design process with special consideration for users experiencing functional limitations. This paper showcases how accessibility concerns affect us all and that a proper definition of accessibility goes beyond individuals with disabilities. We then identify existing accessibility standards and guidelines, and propose design patterns as a proactive medium to capture and disseminate best accessibility design practices. The patterns provide usable solutions to specific accessibility problems, within a defined context, that take into account the physical, cognitive and functional limitations of users. Finally we propose a classification scheme for patterns based on usability principles and accessibility heuristics.

1 Introduction

Since the introduction of the graphical user interface in the 1980s, the ways in which information can be presented on a computer screen have posed challenges to users with functional limitations. For users with disabilities, the challenges have ranged from translating the information to a usable form (a blind person converting screen text to speech output), to manipulating content via input devices.

Accessibility barriers exist on most websites and systems, making them difficult to use for individuals experiencing any functional limitation. To date, most attempts to solve the challenges of the accessibility of digital information have been focused at the technical level. Developers have been encouraged to follow practices which have been found to work better for users with specific disabilities. However, technical solutions to these problems have never been sufficient, largely because the standards and training are always lagging behind technical developments. With the recent proliferation of new mobile platforms and the splintering of the pc-centric form factor, the problem is now more complex.

There needs to be a shift in current thinking. Similar to the adoption of usability practices into the software engineering process, accessibility techniques need to be considered as an integral part of usability, and in turn, the software engineering process. Accessibility concerns should be considered as early as possible starting from the requirements and analysis phase. Errors in requirements are the most costly to repair

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 248–257, 2011.
© Springer-Verlag Berlin Heidelberg 2011

in subsequent development stages [1]. Inconsistencies and ambiguities propagate to downstream design and development where the effort for addressing those concerns multiplies [2]. The likelihood also increases that changes made to address accessibility deficiencies will be put in place in development by workers without design and accessibility expertise. An example would be an ESL developer being asked to write the English alternative text that is deemed necessary for unlabelled images.

By incorporating accessibility considerations into mainstream practices, we should ensure a more usable experience for all users, regardless of functional limitations. Instead of just providing suitable alternatives for disabled users, solutions should address any user experiencing limitations within a usable context.

Accessibility standards, heuristics [3] and guidelines [4, 5] (for web applications in particular) bias towards technical solutions by developers, as opposed to design solutions by information architects [6]. As such, they tend to be reactive solutions to discovered problems, whereas usable solutions require proactive considerations of potential problems [7]. As well, the standards and guidelines are frequently not conducive to rapid adoption or consultation.

In this paper we propose design patterns as a proactive medium to capture and disseminate best accessibility design practices. Patterns alleviate many of the shortcomings associated with guidelines. Patterns only capture essential details of design knowledge in a specific context, and abstract away from superfluous, toolkit-dependent and platform-specific design information. The designer is told when, how and why the pattern can be applied. Design patterns provide usable solutions to specific accessibility problems, within a defined context, that take into account the functional limitations of users.

The remainder of this paper is structured as follows. In the next section we discuss how accessibility applies in a wider context beyond users with disabilities. Section 3 provides a definition of *usable* Accessible Design. In Section 4 we propose patterns for *usable* Accessible Design, provide an illustrative example and propose a classification scheme. Section 5 reviews relevant related work. Finally, we conclude and discuss future avenues.

2 Accessibility Affects Us All

If something is *accessible* it is readily usable. So how does accessibility differ from usability? The term "accessibility" was coined to designate the concept of things being accessible for people with physical or cognitive disabilities. As such, accessibility is sometimes viewed as a subset of usability – employed when a design requirement identifies a need for use by Persons with Disabilities. However the meaning of accessibility can be expanded, especially in the context of modern technology.

A usability architect conceives of a (proto)typical user – a user who represents the "norm" – and designs for such a user's requirements. Although the use of multiple personas [8] can capture specific users that may be considered boundary cases, these still tend to highlight differences of role and task, as opposed to functional limitations. From this standpoint, usability can be said to consider the "representative user" needs. Accessibility, alternatively, is focused on atypical scenarios and users with functional limitations. The field of accessibility's predominant bias is concerned with outliers.

Modern communication mediums have created an abundance of scenarios in which it becomes increasingly difficult to design for *just* a typical user, so many of the principles and guidelines of accessibility can become very relevant to the field of usability. In order to gain a fuller appreciation of how media can affect usability, it is worthwhile looking at some examples.

The mass medium of radio is functionally limited to sound. In essence all users of radio are functionally visually disabled; they cannot use sight (or any of their other senses) to listen to the radio. The design of radio content – all the programs and advertising – is designed around the fact we must use our aural sense. Content is created that seeks to overcome the limitations of the medium: Commentators describe what they are seeing to translate the visual experience to the medium; sound effects are readily employed. A whole series of conventions were developed in the medium to assist the ease of its use (commercial breaks, intros and outros, etc).

All technology media, from silent movies to texting, bias towards one or more of our senses, and usability conventions develop in response. However designing for newer technologies is more complicated because the format of information and the devices for intermediating the information have multiplied. In such a scenario, the percentage of users who experience information in a "normal" context diminishes and the chances that users are in some way functionally limited increases. Whether we try to hear a cell phone call on a busy street or try to read a large data table on a small screen, we all, in essence, experience a functional limitation analogous to the experiences of persons with physical disabilities. The lessons learned from the field of accessibility, then, can be applied in many scenarios to capture the outlier experience, ensuring that things are designed towards universal instead of "typical" use.

This concept of functional limitations can be applied to many contexts. Limitations can be environmental (noisy office, moving vehicle, significant weather), physical and cognitive (what we typically think of as disabilities), cultural (language, age-related) or technological (limitations of the medium or intermediaries, including devices – small screens, limited keyboards, etc).

3 What Is *Usable* Accessible Design?

The overarching goal of *usable* Accessible Design is to design usable software that addresses the potential needs of all users in the user-centered design process, with special consideration for users experiencing any functional limitations.

Usable Accessible Design is concerned with making applications accessible by providing an interface that allows people with functional limitations to perform and complete Information Technology (IT) tasks or use IT content, while ensuring that the user's needs and experiences are met. It is about getting Accessibility and Usability to work together by combining ideas from both fields.

Usability is the extent to which a product enables users to achieve their goals with effectiveness, efficiency and satisfaction in a specified context of use [9]. This definition relates to the quality of the interaction between the person who uses the product to achieve actual work and the product or software application itself, including its user interface.

Accessibility is a quality attribute of a product or service relating to the extent to which it can be used by everyone. When applied within the context of Web and software development, accessibility's basic premise is: That a user's functional limitations should not prevent them from using the technology effectively. Accessibility encompasses all disabilities that affect access to any IT medium, including visual, auditory, physical, speech, cognitive, and neurological disabilities [10]. As well, it includes technical, cultural and environmental limitations such as ESL (English as a second language for non-native speakers) and noisy workplaces.

A bias exists towards using development techniques to resolve existing accessibility issues (sometimes referred to as "technical accessibility"). Such techniques tend to be reactive rather than proactive, with the developers responding to issues as they are discovered by testers. Moreover, the technical accessibility approach favors terminology used by developers that does not match or easily integrate with the vocabulary of designers. Since accessibility issues bridge both disciplines, this can lessen the success with which potential accessibility issues can be proactively flagged and resolved at the wireframe stage.

We advocate a move toward what has been called "Usable Accessibility". By incorporating some of the vocabulary and principles of newer accessibility standards, proven techniques such as personas and design patterns employed in the field of usability can be extended beyond considerations of the representative user to include the specific needs of users or situations which previously may have been viewed as atypical. There are a variety of ways in which these gains can be realized. These can include: Designing content for variable presentation; creating accessible personas; including persons with disabilities in the user-centered design process; and developing (or modifying existing) design patterns. Each of these approaches warrants further discussion and consideration – more than can be provided within the confines of this paper. The next section elaborates on how one of these approaches — design patterns — can be applied to achieve usable accessibility.

4 Patterns for U*sable* Accessible Design

Patterns aim to capture and communicate the best practices of user interface design with a focus on the user's experience and the context of use. As a result, they are an attractive UCD technique, with interesting ramifications for designing across a variety of contexts. Patterns alleviate many of the shortcomings associated with guidelines. Patterns only capture essential details of design knowledge in a specific context, and abstract away from superfluous, toolkit-dependent and platform-specific design information. The designer is told when, how and why the pattern can be applied.

4.1 Extending Patterns to Include Functional Limitations

Every pattern has three necessary elements, which are: A context, a problem, and a solution. The context describes a recurring set of situations in which the pattern can be applied. The problem refers to a set of forces, i.e., goals and constraints, which occur in the context. The solution refers to a design form or a design rule that can be applied to resolve the problem. Other attributes that may be included are additional design rationale, specific examples, and related patterns.

We propose the use of *usable* Accessible Design patterns to extend existing design patterns, by including both accessibility and usability considerations. Similar to other patterns, they provide a solution to a common problem within a particular context of use. However, they also include references to or descriptions of particular *Functional Limitations*. As illustrated in Figure 1, the functional limitation is closely related to the problem of the pattern and its context of use. It also serves as a driver to the solution of the pattern, which must address the functional limitation.

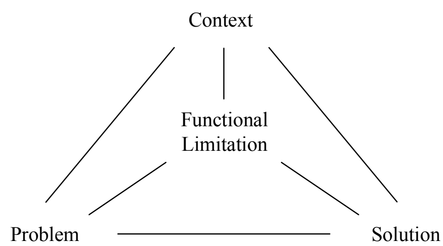


Fig. 1. Structure of an Accessible Design Pattern

An example pattern (*Iconic Navigation Page pattern*) is given in Table 1. The pattern is applicable for users who want to find specific information about a targeted topic, and relates to a variety of functional limitations. The pattern follows the basic structure of Figure 1 and includes an illustrative example. The problem statement of the pattern is immediately followed by a list of functional limitations for which this pattern is applicable. This includes users with low literacy, cognitive limitations, low vision, reduced manual dexterity, using a medium with a reduced screen size, and non-native speakers. Based on these functional limitations, the emerging user needs are as described below. Users to which these needs apply are described in brackets.

- Simplified language (non-native speakers, low literacy, cognitive limitations)
- Simplified interface (cognitive limitations, low vision, reduced screen size)
- Concise language (low vision, reduced screen size)
- Visual reinforcement of meaning (non-native speakers, low literacy, cognitive limitations)
- Validation of choice (non-native speakers, low literacy, cognitive limitations, low vision, reduced manual dexterity)
- Tolerant direct manipulation (low vision, reduced manual dexterity)

One target user group for the Iconic Navigation Page pattern are low literacy users like *Serge* (Table 2), who's functional limitation consists of having a reduced level of reading comprehension. Serge is depicted as a *persona*, which is a user-centered design technique that aims to direct the focus of the development process towards end users and their needs.

Table 1. Iconic Navigation Pattern

Title	**Iconic Navigation Page Pattern**
Problem	How to reduce the complexity of multi-level navigation and simplify access to navigation without necessarily impeding access to the content.
Functional Limitations	- Low Literacy: users who have a reduced level of reading comprehension - Cognitive limitations: users with reading, linguistic or visual comprehension deficiencies - Low Vision: users who require enlargement or magnification of content to see it properly - Reduced manual dexterity: users who have a physical impairment that prevents easy operation of traditional input devices - Reduce screen size: Users whose ability to see content is limited by the physical size of their device - Non-native speakers: users whose first language is not the same as the content of the web site
Context	Traditional site navigation is not optimal for some users, due to functional limitations. The *Iconic Navigation Page* pattern is an alternative to traditional site navigation, used to simplify navigation and associated content.
Solution	Using a minimalistic approach, this pattern reduces the complexity of the navigation architecture by selecting the most important elements/functions and presenting them to the user in a simplified interface. By separating the navigation architecture from the resulting content, the complexity is reduced both for the task of navigation and for the task of reading primary content. Hierarchical nesting of navigation items should be kept to a minimum in order to limit the cognitive load. Icons that visually reinforce the meaning of the choices are provided, along with labels. The large icons, separated by white space, provide a large clickable area, providing tolerant direct manipulation for users who have difficulties make making precise movements with input devices such as mice.
Examples	 Use of Navigation Page Pattern in [11]

Personas describe target users of a system, providing stakeholders and developers with a better understanding of both the users and context of use for a planned tool or interactive system. They encapsulate behavioral data [8] gathered from ethnography and empirical analysis of actual users, in addition to demographic information, needs and preferences. A typical usage scenario including interaction details for Serge could be as follows:

Serge is home for the evening, and wants to find some information on the municipal web site about the Montreal Access Card. His co-worker told him that the card can give him discounts on a number of services in Montreal, including recreational and sports-related entrance fees, and that he can find all the information online. He wants to click on a link which he thinks will lead him to the right page, but he is unsure. He hesitates for a few seconds, and then decides to click. He tries to make sense of the sentences on the page, but there is way too much information. He isn't sure where to click next. He gets tired, and calls it a night, deciding that maybe he will try again tomorrow.

Personas can be used to identify user needs, giving insight into a user's functional limitations, which in turn have implications on design considerations and constraints. Therefore, they could be thought of as a natural starting point in identifying appropriate patterns that are classified according to usability principles and accessibility heuristics (discussed in the next section). Because patterns are a common UCD tool, the incorporation of functional limitations into existing patterns can easily embed accessibility considerations into the process.

4.2 Towards a Pattern Language for *usable* Accessible Design

Pattern languages are a structured method of describing best design practices and are a means to traverse common usability problems in a logical way, describing the key characteristics of effective solutions for meeting various design goals [12]. To effectively take advantage of pattern languages, a logical grouping of their entailed patterns is required. We propose a two-dimensional classification scheme in which patterns relate to (1) a set of usability principles and (2) a set of accessibility heuristics.

For example, the Iconic Navigation Page patterns relates to the following two usability principles defined in [12]: *Facilitated Navigation* and *Minimalistic Design*. The former refers to the means used to facilitate "movement" through the contents of an interactive program in an intentional manner whereas the latter refers to the use of only required items in order to create a clean and aesthetic design. The pattern however, also relates to two orthogonal accessibility heuristics defined in [3]: *Tolerant Direct Manipulation* and *Logical Presentation*. The former allows for imprecise and/or slow pointer movements so as to assist users with low vision or physical impairments such as shaky hand movements having difficulty using smaller targets. The latter refers to characteristics of objects and actions that should facilitate hierarchical, sequential or simple grouping associations so as to assist users with cognitive limitations who may have difficulty discovering the organization of more complex elements and hierarchical structures.

By using this classification scheme of patterns it is possible to capture both essential usability and accessibility principles as part of the pattern language. Moreover, it allows for traceability, which allows one to trace whether a given pattern addresses a certain heuristic and vice versa.

Table 2. Sample Persona "Serge Lacroix"

Serge Lacroix	 **"I want to be able to find information online without having to ask my son for help"**
General Profile	Serge is a 51-year old city worker. He lives with his wife of 30 years. He has three grown children and two grand-children. He is a blue-collar worker, and has been working for the city for over 25 years. He is worried about his retirement, and is careful with his spending habits. He has a decreased ability to read and write, although he is more comfortable with using numbers. For public service inquiries, he usually makes telephone calls. For figuring out banking and billing information, he gets help from his son. Through a new initiative with the city of Montreal, he heard about easy-access information on some websites, which is now supporting users like him. He is nervous about dealing with the internet, but he would prefer this over asking his son all the time for help
Goals	Professional: Maintain his job until retirement. Personal: Increase his savings, spend time with his grand-children. Application: Find basic information.
Demographics	Serge is a 51-year old male, with somewhat of a low family income. He has two financial dependents (his wife and mother).
Knowledge and Experience	Serge is a native French speaker. He has a low literacy level, and did not finish elementary school. He has little experience with computers and the internet.
Psychological profile and needs	He has an extremely high need for validation of decisions when interacting with a computer system. He is somewhat of a reactive user, and has a low learning speed. He also has a somewhat high need for guidance during system interaction.
Attitude and Motivation	He has a negative attitude to IT, but an average level of motivation to use the system.
Functional Limitation	He belongs to a special user group, *low literacy*.

5 Related Work

The Web Content Accessibility Guidelines (WCAG) [4], published by the World Wide Web Consortium, are now in their second version and are the most widely adopted and recognized accessibility standards in the world. Nonetheless, they pose their own challenges for someone trying to test against or be guided by them [6]. Although there has been an effort to use simple, clear language, the resulting guidelines – and even their structure – can be somewhat impenetrable. The terms given to categorize issues are not always clear, and the way the categories break down is not always intuitive. This is particularly true of the area covered by each guideline and its criteria. The guidelines follow a hierarchy, each level increasing the complexity by a factor of 3 or 4. There are four principles, 12 guidelines, 38 success criteria (at level AA), and 100s of techniques and common failures.

As an antidote to the complexity of such guidelines, some have advocated for the use of accessibility heuristics, which can more briefly address many common shortcomings of design while ensuring a broader range of functional limitations are considered [6]. By incorporating accessibility heuristics into a common usability evaluation technique, the considerations of accessibility are more effectively brought into the user-centered design process. Such heuristics are more precise and succinct than guidelines: IBM's 11 usable access heuristics [3], along with descriptive examples, easily fit on one page. Although these were considered incomplete and a work in progress when published in March 2009, it is easy to see why they have value for UI design. The language is more easily grasped and more memorable [3]. A primary challenge to heuristics is they are most effective as reinforcement for existing expert knowledge. Like guidelines, at a fundamental level they are only as effect as the person employing them. The user still requires accessibility education. A second challenge is that while they will certainly lead to websites that are more usable for all, they will not necessarily conform to legislated guidelines.

Another way of simplifying the adoption of WCAG guidelines and techniques is through the use of a mitigating design pattern language. In [7], Folgi et al. propose a design pattern language for accessibility. The approach maintains the hierarchical breakdown of the W3C standards, but seeks to provide it within a context more familiar to usability architects. Because the authors base their model on the WCAG standards, they inherit some of the same problems with normalization. Thus, their model presents three broad, high-level categories – Multi-formal presentation, Easy Navigation and Interaction at user's pace – but items in mid- and low-level categories can fall under more than one high-level category. Moreover, the low level patterns are analogous to techniques and are biased towards developer practices.

The pattern classification scheme presented in this paper is in stark contrast to the approach by [7]. Instead of hierarchical nesting we propose classifying patterns according to existing usability principles and accessibility heuristics. We believe that such an approach provides a more natural mapping and it allows for more flexibility as the constraint of grouping all patterns under a common root pattern is removed.

6 Conclusion

Usable Accessible Design is about designing usable software that addresses the potential needs of all users in the user-centered design process – with special consideration for users experiencing functional limitations. In this paper we broke with the predominant assumption that relates accessibility exclusively with persons with disabilities and argued that accessibility concerns affect us all. Our wider definition of accessibility incorporates all sorts of functional limitations including physical, cognitive, technological and environmental.

We argued that existing accessibility standards, heuristics and principles, for web applications in particular, bias towards technical solutions. These approaches attempt to address accessibility near the end of the software development lifecycle by identifying shortcomings of a near-completion project. As such, they tend to be directed at developers and testers, encouraging a reactive rather than proactive approach. In order to overcome the bias, we proposed incorporating accessibility concerns throughout the development lifecycle starting with the requirements phase using a pattern-oriented design process.

We presented patterns for *usable* Accessible Design which incorporate functional limitations as an additional pattern component. These functional limitations closely relate to the problem of the pattern and its context of use. It also serves as a driver to the solution of the pattern, which must address the functional limitation. As an illustrating example, we presented the *Iconic Navigation Page* pattern used to reduce the complexity of traditional navigation and is especially suited for users with, for example, cognitive limitations, or users working in mobile environments.

Finally, we proposed a classification scheme for patterns based on usability principles and accessibility heuristics. Based on the user's needs and functional limitations, we can select patterns that are classified according to usability principles and accessibility heuristics. Because patterns are a common UCD tool, the incorporation of functional limitations into existing patterns can easily embed accessibility considerations into the process. We believe that such a classification simplifies the task of identifying an appropriate pattern. As for future work, we plan to further refine the classification scheme and propose a systematic process for selecting and applying *usable* Accessible Design patterns.

References

1. Leffingwell, D., Widrig, D.: Managing software requirements: a use case approach. Addison-Wesley, Boston (2003)
2. Boehm, B.W.: Software Engineering Economics. Prentice-Hall, Englewood Cliffs (1981)
3. King, M.: Closing the Competitive Productivity Gap for People with Disabilities: Software Engineering Guidance and Challenges, IBM White Paper (2009), `http://www-03.ibm.com/able/news/accessorusable.html` (last accessed February 2011)
4. WCAG, Web Content Accessibility Guidelines (WCAG), W3C Recommendation (2009), `http://www.w3.org/TR/WCAG20` (last accessed February 2011)
5. WAI-ARIA, WAI-ARIA 1.0 Primer: An introduction to rich Internet application accessibility challenges and solutions, World Wide Web Consortium (2009), `http://www.w3.org/TR/wai-aria-primer/` (last accessed February 2011)
6. Paddison, C., Englefield, P.: Applying Heuristics to Perform a Rigorous Accessibility Inspection in a Commercial Context. In: Proc. of Conference on Universal Usability, Vancouver, Canada, pp. 126–133 (2003)
7. Fogli, D., Parasiliti, L., Bernareggi, C.: A Design Pattern Language for Accessible Web Sites. In: Proc. of AVI 2010, Rome, Italy, pp. 307–310 (2010)
8. Cooper, A.: The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity. SAMS Publishing, Indianapolis (1999)
9. ISO 9241, Standard on Guidance on Usability (1991)
10. Henri, L.S.: Just Ask: Integrating Accessibility Throughout Design. LuLu Press (2007)
11. Ville de Montreal Homepage (2006), `http://ville.montreal.qc.ca` (accessed October 2006)
12. Javahery, H.: Pattern-Oriented UI Design Based on User Experiences: A Method Supported by Empirical Evidence, PhD Thesis in the Department of Computer Science and Software Engineering, Concordia University, Montreal (2006)

From Structural Analysis to Scenarios and Patterns for Knowledge Sharing Applications

Claus Kaelber[1] and Christian Märtin[2]

[1] teamwissen
Entenbachstraße 35
81541 Munich, Germany
kaelber@teamwissen.de
[2] Augsburg University of Applied Sciences
Faculty of Computer Science
Friedberger Straße 2a
86161 Augsburg, Germany
Christian.Maertin@hs-augsburg.de

Abstract. In this paper we present a pragmatic development approach for knowledge sharing applications that encompasses both design and software engineering aspects. It starts from scenarios and leads to patterns that help application developers and user interface designers on the one hand to separate relevant content from unimportant data and on the other hand propose techniques for qualitatively structuring knowledge management and knowledge sharing tasks for enterprises and individuals.

Keywords: HCI patterns, structural patterns, domain patterns, knowledge sharing, knowledge management, design strategy, GUI generation, pattern-based modeling.

1 Introduction

A steadily growing information flood faces people both in business and everyday life. The many existing commercial solutions for knowledge management and knowledge sharing systems, however, in most cases even raise the complexity level for accessing relevant information or come with poorly designed user interfaces that require experts and keep other users away from working with the systems. We present a new approach starting with a requirements analysis driven by contextual design techniques that leads to scenarios from which structural, domain and HCI patterns are derived.

These patterns can be exploited by tools and generators in order to arrive at high-quality web-based user interfaces for knowledge sharing applications. These knowledge sharing applications are aimed at optimizing user experience and usability in contrast to powerful, but complex existing knowledge sharing environments and workflow management systems.

The paper discusses the final results of project *p.i.t.c.h.* (*p*attern-based *i*nteractive *t*ools for improved *c*ommunication *h*abits in knowledge transfers) which was conducted by the *A*utomation in *U*sability *E*ngineering group *(AUE)* at Augsburg

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 258–267, 2011.
© Springer-Verlag Berlin Heidelberg 2011

University of Applied Sciences in cooperation with two medium-sized enterprises with engineering and production background and several partners from communication sciences and the knowledge management domain.

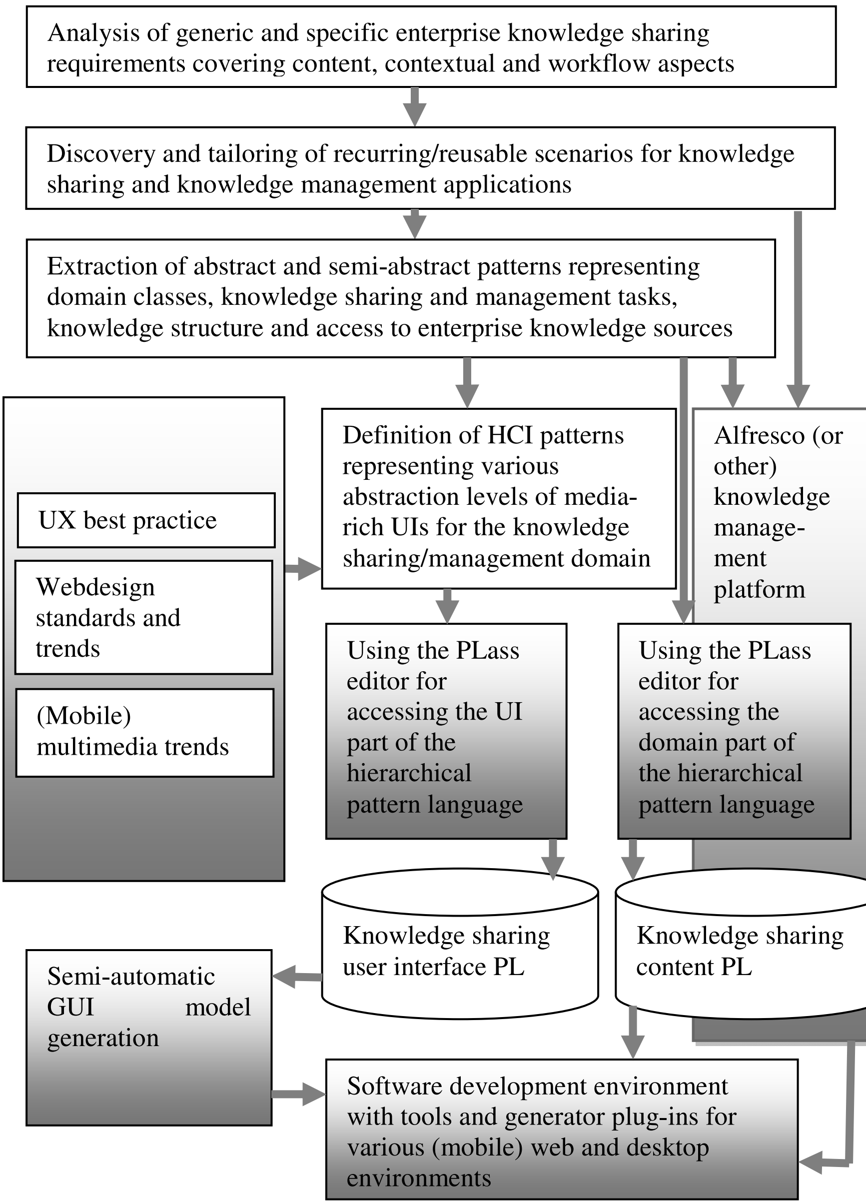


Fig. 1. Revised p.i.t.c.h. development workflow

The p.i.t.c.h. development environment is described in [6]. Figure 1 shows a revised version of the application design workflow that was devised in p.i.t.c.h. It illustrates how the content-specific analysis phase interacts with our scenario- and pattern-based development environment.

This paper focuses on the scenario-based modeling and design process that also includes techniques from contextual design [1], the resulting patterns and the final translation into interactive applications.

2 Design Strategy

One of the most important goals of the p.i.t.c.h. project was to set up a design strategy that would lead to more intuitive, simple functionality for searching, finding, storing and accessing data. See [4], [5] and [8] for background information on knowledge management systems. An appropriate strategy should enable the extension of existing content and the exchange of content between users in the same or in different organizations. It should also identify the individual skill-levels of the various users and their topics of interest.

Therefore the central requirements for the design strategy in p.i.t.c.h were

- reduction in knowledge sharing system complexity
- intuitive system usage
- fast system access
- interactive knowledge access and communication tools
- individual system configuration

In order to arrive at the required goals the project was structured into the following phases

- documentation and analysis of existing IT- and communication workflows in the participating enterprises
- definition of a design rationale for the flexible integration of multi-media formats into the target system
- definition of typical lead-scenarios
- derivation of reusable HCI patterns from requirements and scenarios
- tool-based automated development and test of interactive target applications

The aspects of joy of use and of rich user experience are still far away from being considered by most application developers, because they are mainly domain-focused or driven by application functionality. These rather psychologic properties of digital media access can be exploited to add substantial value to knowledge sharing environments. Driven by the prevalence of social networks future methods of information brokerage within organizations will need more collaborative and personalized communication tools. Therefore design strategies and development processes used in these application domains must provide a (semi-)formal basis for also including such rather informal user experience and personalization aspects into the development work.

Using scenarios that lead to patterns, enabling the patterns to also express UX [11] and non-formal individualization aspects, and using semi-automatic tools for generating the interactive parts of the final target applications was the key to the design strategy applied by the p.i.t.c.h. project.

In order to evaluate the design strategy with a prototype in a demanding and useful target context, during the initial project phase typical knowledge sharing and knowledge communication requirements of three medium-sized technology-driven enterprises were analyzed using contextual design inquiry techniques. The discovered requirements were clustered and scenarios that cover the most important workflow, knowledge access and communication aspects were derived.

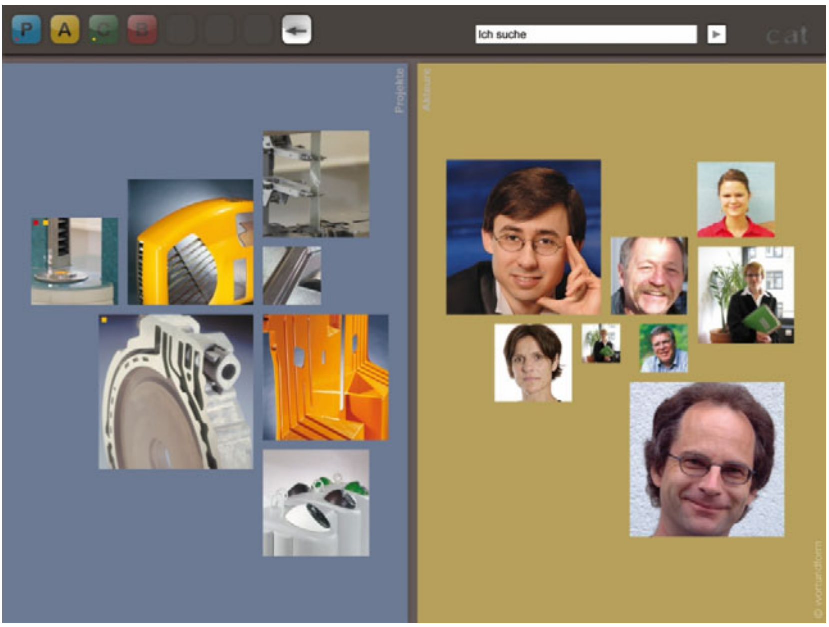


Fig. 2. p.i.t.c.h. orientation environment

From these the following lead-scenarios were selected for the prototype and as the basis for extracting reusable abstract, semi-abstract and concrete HCI patterns:

- Profile Board
- Infolog
- Collaborative Reporting

The scenarios apply the same basic organizational scheme that uses the *login* pattern to individually access initial *orientation environments* that allow for a global overview and further orientation within the collaborative knowledge sharing system. The starting point is represented by three abstract structural patterns: *orientation environment, protagonists* and *projects*. These patterns provide the infrastructure for accessing all relevant information and serve as the pivotal structures for all collaborative and communicative actions of the users.

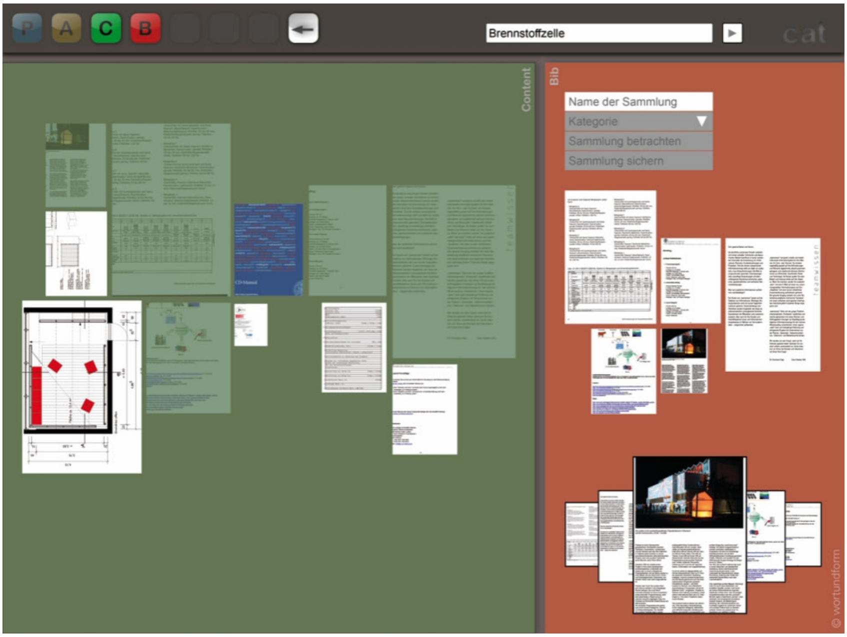


Fig. 3. User interface generated from document library pattern, domain and HCI patterns

The orientation environments are individually personalized (fig. 2). They are characterized by a specific visual ranking of the relevant content data (their size, color, position depending on up-to-dateness, scope, deadline etc.). Patterns were developed to represent these mappings. In the figure the individual network of protagonists and projects is visualized in the form of an information cloud or map. All functionality, internal, external links, documentation and communication tools are accessible from these orientation environments. Figure 3 shows the document library for a specific project. HCI patterns [7], containing the mappings from the XML content models to semi-abstract UI representations, were modeled for various types of knowledge sharing and communication tasks.

3 HCI Patterns and Content Models

Patterns in our environment are specified interactively using the PLass editor [12]. They are modeled with similar attributes, e.g. *name, problem, forces, context, solution, examples,* as can typically be found in design and HCI pattern languages (e.g. [2], [3], [9], see [13] for a comprehensive overview on pattern-based development of interactive systems). Internally they are represented in XML notation.

In addition to classic design patterns, our patterns provide the powerful attribute *automation* that either provides structural, content, and contextual information for refining more abstract into more concrete patterns or provides linking information from a concrete pattern to a single or a group of concrete interaction objects that can directly be exploited by automated user interface generator plug-ins.

XML is also used for content and domain modeling of the chosen application types. The *Alfresco* CMS [10] in its version 3.4 is used as our underlying model repository and knowledge server.

Starting from the structural, functional and workflow requirements that were defined during analysis and led to the resulting prototypical scenarios, both HCI and domain patterns could be derived from the scenarios. The domain patterns mainly define the domain classes, their attributes, functional and communication behavior. In order to allow an easy mapping to the target environment, the content of the *solution* attributes of the domain classes was directly mapped to XML representations of each domain class in the content model residing in the Alfresco model repository.

Alfresco offers useful open interfaces (e.g. RESTful Services (REST), Alfresco Web Scripts) for coupling and synchronizing data-oriented servers with visual clients and supports both Javascript and Java for developing the interactive client behavior. Alfresco´s REST provides an URL-based interface for resource access and state transfer between server and clients. It is used to access content data stored in the knowledge repository and map them to the final UI structure that was derived from the specified scenarios and HCI patterns.

Alfresco Web Scripts apply the MVC pattern with the XML based content model residing in the Alfresco repository, Java or Javascript as the Web Script controller language and the Freemarker Template Language as the toolkit for providing the visual elements (view). In order to support different target environments plug-ins for HTML 5, Javascript, Java, Adobe Flash, and Ajax were used by the interactive clients. The HCI patterns that were identified during analysis and scenario definition were linked to the representations of the domain patterns, kept in the model repository.

Figure 4 shows the structure and relationships of the various patterns and model components defined for the prototypical application based on the three example scenarios. The figure focuses on one of the scenarios, the *profile board.*

A scenario can itself be seen as a pattern that specifies the use-cases, possible tasks, workflow, domain knowledge as well as the structural and behavioral requirements of the user interface. Resulting from initial requirements analysis and an extensive contextual inquiry phase, the scenario-pattern attributes were specified as structured texts. Scenario patterns serve as the most abstract elements in the project pattern language. To link them to the more formal patterns of the later development stages the scenario-patterns are entered directly below the root element of the pattern language. The pattern attribute *automation* is filled with links to all structural patterns that are directly referenced by the scenario pattern.

From the beginning the project team included domain experts, designers, and computer scientists. Therefore, whenever possible, the most important structural patterns were already identified at this stage.

For the profile board scenario the structural patterns *orientation environment, protagonists, and projects* were identified (dashed lines indicate that a scenario uses these patterns). These patterns are reusable for model specifications throughout the entire knowledge sharing and knowledge management domain. The structural patterns are specified using the regular (HCI) pattern attributes provided by the PLass editor. As can be seen in fig. 4, structural patterns may reference other structural patterns, e.g. the *projects* and *protagonists* patterns are used by *orientation environment.* In the resulting pattern language *orientation environment* appears at a higher level in the pattern hierarchy.

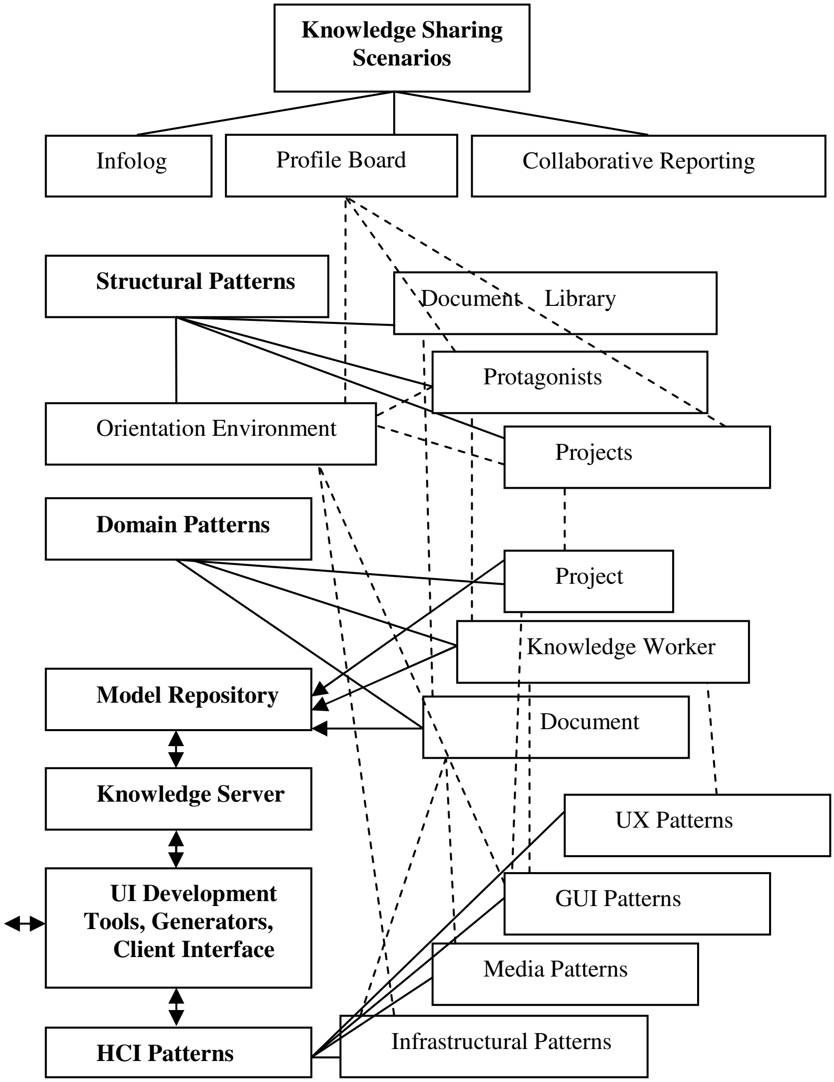


Fig. 4. Excerpt from the p.i.t.c.h. model structure

At the center of the knowledge sharing applications developed by the project a model of the major domain classes and their features is maintained in the model repository.

However, the domain classes are directly derived from the domain patterns identified during the analysis process. Domain patterns like e.g. *knowledge worker* or *project* specify the data structure of single or composed domain classes, their functionality and inter-class communication requirements. Domain pattern attributes are specified as structured text. Their *automation* attribute contains the XML specification data and functionality that is stored by the Alfresco-based model repository during runtime together with some glue code. Thus the mapping of domain patterns to the underlying CMS (see the arrowed lines) can be automated. For the final application, the respective enterprise data can be filled in using the CMS templates structured after the domain pattern specifications.

As can be seen from fig. 4 domain patterns are separated from, but interact with the HCI patterns that specify all aspects of GUI representation, media requirements, contextual mapping to varying target environments as well as user experience. Using the automation mechanisms and plug-ins provided by the *PAGui* tools [12], parts of the GUI and interactive behavior of the target application can be generated, other parts are implemented with the HTML 5, Java and Javascript-based Web-development tools supported by Alfresco.

Table 1. User experience pattern for visually arranging project contributors

Attribute	Description
ID	Arrange protagonists
Problem	How to visually arrange protagonists that contribute to a project
Context	Find the most relevant communication partners when working on the project details
Solution	1. Compare the following attributes of the protagonists: last contact, type of contact, number of contacts 2. Evaluate the results in a layout algorithm that returns a position, size and coordinate-pair for each protagonist image. The algorithm is running each time the detailed project view is activated.
Illustration	Refer to figure 5.

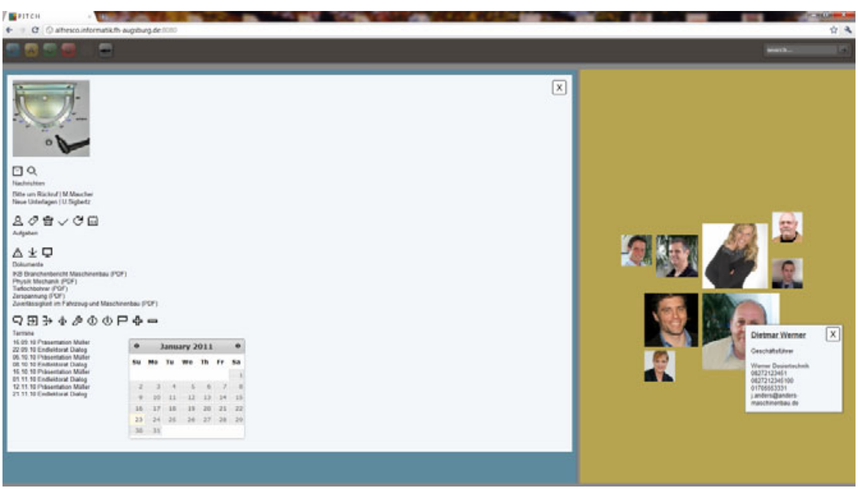


Fig. 5. User interface generated from *project*, *protagonist*, GUI and UX patterns

In figure 5 a partly generated user interface is shown that was derived from the structural patterns *project* and *protagonist*. It shows the detailed view of a project with documents, a to-do-list and a calendar. These properties are modeled in the *project* pattern. On the right-hand side, the project contributors are visualized by their images. A UX pattern (see table 1) is exploited for generating the layout of the contributors´ images.

4 Conclusion

The development workflow and the design strategy defined within the p.i.t.c.h. project were used to design and implement several prototypical knowledge sharing scenarios that serve as the basis for future, more comprehensive knowledge sharing environments for small and medium sized enterprises. The pattern language developed by the project is reusable and can be extended for additional domains.

The *profile board* scenario introduces protagonists with all their projects, fields of interest and their individual business network, whereas the *infolog* scenario presents an aggregation of media-dependent personal or project-specific content data in the form of an individualized, interactive kiosk. The *collaborative reporting* scenario establishes separated rooms for communicating and exchanging information with external business partners. These scenarios currently serve as an experimental platform for evaluating the quality and user experience possible with our underlying design strategy and the pattern-driven development process.

Future development stages will focus on mobile access to knowledge sharing applications, advanced visualization concepts for project-related real-time workflows and high-level automation tools for the development process.

Acknowledgements. Part of this research was supported by the Innovation Fund of the IHK Schwaben (Augsburg Chamber of Commerce), Augsburg, Germany. Our thanks also go to Christian Herdin and the student members of the M.Sc. CS project in the winter semester 2010/11 who were responsible for implementing the concrete UI patterns, the Alfresco content models and the domain part of the sample scenarios.

References

1. Beyer, H., Holtzblatt, K.: Contextual Design. Interactions, 32–42 (January-February 1999)
2. Buschmann, F., et al.: Pattern-Oriented Software Architecture – A System of Patterns. Wiley, New York (1996)
3. Designinginterfaces.com, Jennifer Tidwell, Patterns for Effective Interaction Design (March 10, 2010) , `http://www.designinginterfaces.com/`
4. Roumois, H.: Ursula: Studienbuch Wissensmanagement, Zürich, p. 34 (2007)
5. Lehner, F.: Wissensmanagement, Munich, p. 46 (2008)
6. Märtin, C., Engel, J., Kaelber, C., Werner, I.: Using HCI-patterns for modeling and design of knowledge sharing systems. In: Forbrig, P., Günther, H. (eds.) BIR 2010. LNBIP, vol. 64, pp. 1–13. Springer, Heidelberg (2010)
7. Marcus, A.: Patterns within Patterns. Interactions, 28–34, (March-April 2004)

8. Trondsen, E., Vickery, K.: Learning on Demand. Journal of Knowledge, Management, 169 (March 1998)
9. van Welie, M.: Welie.com, Patterns in Interaction Design (March 10, 2010), http://www.welie.com/patterns/
10. Alfresco ECM System product homepage (March 10, 2010), http://www.alfresco.com
11. Obrist, M., et al.: User Experience (UX) Patterns for Audio-Visual Networked Applications: Inspiration for Design. In: Proc. NordiCHI 2010, October 16-20, pp. 343–352 (2010)
12. Engel, J., Märtin, C., Forbrig, P.: Tool-support for Pattern-based Generation of User Interfaces. In: PEICS 2010, Berlin, Germany, June 20. ACM Int. Conf. Proc. Series, pp. 24–27 (2010)
13. Breiner, K., et al. (eds.) Proc. of the 1st Int. Workshop on Pattern-Driven Engineering of Interactive Computing Systems (PEICS 2010), Berlin, Germany, June 20. ACM Int. Conf. Proc. Series (2010)

A Design Patterns Approach to Adaptive User Interfaces for Users with Special Needs*

Matthias Peissner, Andreas Schuller, and Dieter Spath

Fraunhofer Institute for Industrial Engineering (IAO),
Nobelstr. 12, 70569 Stuttgart, Germany
{matthias.peissner,andreas.schuller,
dieter.spath}@iao.fraunhofer.de

Abstract. Providing truly accessible user interfaces for a great variety of users still presents a major challenge for software developers. Contrasting from a 'design for all' approach, we present a design patterns based approach for the implementation of adaptive user interfaces for users with special needs. This paper gives an overview of essential requirements of adaptive systems, covering aspects for users as well as system developers. The overall structure, adaption mechanisms, distinctive design pattern types and demonstrative examples of the 'MyUI' design patterns approach are described and illustrated.

Keywords: Adaptive user interface, design patterns, accessibility.

1 Introduction

Adaptive user interface technologies have a long tradition (see e.g. [1]) but still are envisioned to yield great advantages for the usability and accessibility of advanced technical systems (cf. [2]), especially for people with special needs. Modern recognition, sensor and agent technologies aim at interpreting user actions and contextual information in terms of user goals and intentions and providing services as an adequate system reaction. Adaptive user interfaces working on user and context models are widely considered as empowering the user to benefit from system intelligence and to control complexity.

However, most of the published approaches to adaptive user interfaces aim at different purposes than accessibility for users with special needs. Early systems as the Doppelgänger (cf. [3]) and the Lifestyle Finder (cf. [4]) were built in order to present or suggest interesting content to a user with specific interests. In their recent book, Jannach et al. [5] provide an extensive overview of personalized recommendation systems in diverse application fields. Similarly, adaptive e-learning systems such as ELM-ART II (cf. [6]) aim at promoting efficient learning by adapting the instruction material to the learner's individual progress and cognitive level.

Another category of adaptive user interfaces support the users by providing appropriate levels of assistance and guidance in order to improve the usability and comfort of use, e.g. "Augur" ([7]), "Flexcel" ([8]) and "SmartCal" ([9]).

* This work was funded by the European Union MyUI project under grant FP7-ICT-248606.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 268–277, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Modern Design-for-All approaches discuss flexible user interfaces which are generated on the basis of a specific user profile, as one common user interface for all often fails to meet many users' special needs. In the EU-project ASK-IT (IST-2003-511298: Ambient Intelligence System of Agents for Knowledge-based and Integrated Services for Mobility Impaired users), a user taxonomy has been developed which can be used as a starting point to describe specific needs (cf. [10]). The follow-up project OASIS has developed an adaptation framework which relies particularly on a library of adaptive widgets (cf. [11]). In the research project SUPPLE at the University of Washington, a UI toolkit has been developed for building applications with automatically generated and personalizable user interfaces. Main research and development activities of the SUPPLE project include the adaptation to a person's device ([12]) and physical disabilities ([13]).

According to Brusilovsky ([14]), user interface adaptations can take place in three areas of a system: (1) the selection of *content* to be displayed or recommended to the user, (2) the *presentation* of information including colors, font sizes, layout, etc., and (3) *navigation* which defines the possible paths users can take through an application in order to access a certain information or functionality. In many cases, user interface adaptations concentrate on very specific areas of a user interface whereas other design areas are kept stable. In personalized e-commerce applications, for example, only the selection of displayed content elements is adapted to fit the user's assumed preferences. Adaptive user interfaces which aim at improving the accessibility, however, will have to take into consideration all three design areas. The great diversity in perceptual, motor and physical capabilities and in contextual conditions requires adaptations of content, presentation and navigation. Moreover, there are significant interaction effects between adaptations in the different design areas. Adaptive mechanisms designed in order to improve one aspect of interaction will often lead to increased effort or even problems in another dimension (cf. [15]). For example, when font sizes are increased in order to address vision impairments, selection lists might need to be split to be displayed on two screens, which increases attention and memory demands on the user. Thus, developing user interfaces that support accessibility by providing extensive adaptivity is a highly complex challenge which appears to be not yet mastered.

The current paper presents a design patterns based approach to adaptive user interfaces which is under development in the ongoing European project MyUI (FP7-ICT-2009-4-248606). The project will produce an infrastructure for accessible user interfaces which self-adapt during run-time in order to cover individual needs and limitations. The MyUI adaptation relies on a repository of user interface design patterns which are selected and put together according to the current user profile. A major objective is to reduce the need for configuration or user enrolment. The user profile shall be produced and refined iteratively on the basis of information gathered during the interaction between a user and a MyUI product or service. This paper focuses on the conceptual user interface adaptation framework of the MyUI project, whereas other aspects are covered in further papers of this conference proceedings: [16] discussing the human factors perspective, [17] describing the design of the middleware architecture and [18] presenting the MyUI "Virtual User Lab" which employs virtual reality tools to support designers and developers through simulation of end user interactions in realistic scenarios.

2 Requirements on Adaptive User Interfaces to Improve Accessibility

As pointed out by [16], adaptive user interfaces can help to support accessibility. From a user perspective, the design of effective and acceptable adaptations, however, is not straightforward. A basic challenge is that “the benefits of correct adaptations must outweigh the costs, or usability side effects, of incorrect adaptations” ([15]). This demand can be understood in manifold ways, as, for example, adaptive systems are often criticized for their lack in consistency and therefore dependability and for their autonomous adaptation behaviour which can result in a loss of user control. Major challenges in the design of concrete adaptive user interface solutions include creating design solutions that fit the individual users’ needs and finding adaptation mechanisms that lead to understandable and trustworthy system behaviour. Moreover, the basic principles of adaptation play an important role in deciding upon the effectiveness and success of adaptive user interfaces. For the design of the conceptual framework of MyUI adaptive user interfaces, we focus particularly on the following major requirements:

- **Adapt content, navigation and presentation**
  The MyUI project addresses accessibility problems typically associated with aging and stroke in quite a broad sense. Therefore, MyUI user profiles cover perceptual, cognitive, and motor characteristics and impairments of individual users. As a consequence, all of the three above mentioned user interface domains content, navigation and presentation must be subject to adaptation.
- **Modular and extensible**
  Individualized user interfaces for heterogeneous user groups require a just as broad spectrum of user interface solutions. A modular approach will be needed to manage such a huge amount of possible user profiles and respective user interface solutions. For practical reasons, extensibility of the modular approach will be important in order to support a quick start with a manageable subset of design solutions and later extensions.
- **Learn and adapt during run-time**
  Intelligent user interface adaptation mechanisms avoid tedious configuration procedures. Not only before but during the interaction, the system is collecting information about a specific user which is stored in a user profile as basis for automatic user interface individualizations. This self-learning process leads to dynamic changes of an individual user profile during the interaction that will be reflected in a dynamically adapting user interface. Furthermore, aging and recovering from a stroke are often associated with significantly altering capabilities which will lead to different needs and changed user profiles over time. Therefore, the adaptation framework must support run-time rendering and run-time adaptations of the user interface (cf. “adaptation during use” as described by [19]). The design of transitions from one instance of a user interface to another becomes an important design issue.
- **Consistent and meaningful user interfaces**
  Adapting user interfaces to multi-dimensional user profiles can be viewed as resolving a multidimensional problem where conflicts and inconsistencies can

easily occur. Effective rules and mechanisms for the adequate selection and composition of design solutions throughout an entire application will be fundamental for achieving a consistent and meaningful user interface.

3 The MyUI Design Patterns Approach to Adaptive User Interfaces

3.1 Self-learning and Self-adaptive User Interfaces

MyUI aims at minimizing the need for an initial user interface configuration or user enrolment. MyUI systems are delivered in a "raw" state that have the potential of evolving diverse concrete shapes in combination with a specific end user and her/his surroundings. Then, the system is learning to more accurately adapt to a specific end user and relevant contexts or situations. This adaptation requires user and context information which is shared across all adaptive applications in the environment. The gained synergies result in a network of personal applications which can learn from each other on-the-fly and act in coordination. Figure 1 illustrates the closed loop of interaction, sensing, interpretation, user profiling and composing an individual user interface on the basis of the currently available user profile.

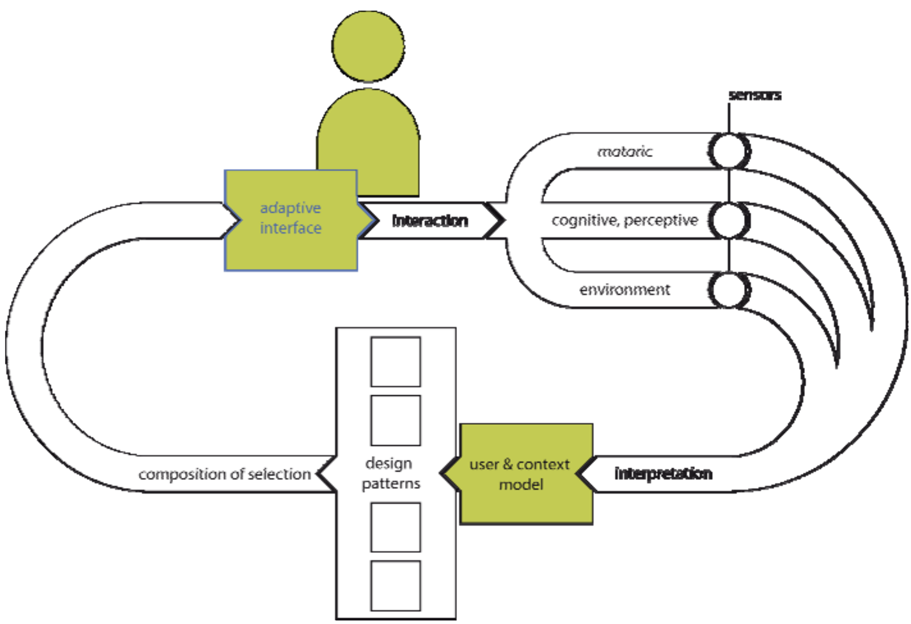


Fig. 1. Self-learning and self-adaptive user interfaces in MyUI

3.2 Design Patterns as Building Blocks of a Modular User Interfaces

In order cover the great heterogeneity of users, MyUI follows a modular approach to user interface development which relies on the composition of multimodal user interface design patterns. The design patterns provide proven solutions for specific interaction contexts, specific end user characteristics and contextual requirements which are explicitly addressed by the design. Each pattern is associated with software components that can be composed and modified in order to achieve accessible user interfaces. Thus, individual accessibility is achieved by combining design patterns that suit for a certain end user and context of use. User interface adaptation is done by switching from one instance of a design patterns cluster (e.g. all patterns for single selection from a list of options) to another design pattern of the same cluster which is hypothesized to fit best to the individual user's needs.

The conceptual back bone of the MyUI is the design patterns repository which includes proven design solutions for optimal accessibility and usability. Each pattern is described in a defined structure as proposed in [20] and related to other patterns of different types and levels of abstraction.

3.3 MyUI Adaptation Framework

The MyUI design patterns repository includes four categories of design patterns:

1. generic patterns,
2. interaction patterns,
3. common patterns and
4. transition patterns.

Each pattern type fulfils distinct functions in the MyUI adaptation framework and therefore, requires a specific description format. *Generic design patterns* define global user interface settings to fit specific user needs and context conditions. On the basis of these global user interface settings, *interaction design patterns* provide suitable user interface elements and widgets for current interaction situations. *Common patterns* define all the stable features of a user interface for a specific device which are not subject to adaptations, e.g. the resolution of the screen, a basic layout grid, etc. *Transition patterns* cover the dynamics of the adaptation processes, i.e. they define the mechanisms of switching from one generic or interaction pattern to another. The range of potential transition patterns goes from requesting explicit user confirmations before adapting the user interface to automatic and sudden changes without directed user action. The design and selection of suitable transition patterns will be crucial for the usability, understandability and acceptability of adaptive user interfaces. Changes in different areas of the user interface will require different levels of user awareness and user initiative in order to be effective.

Generic and interaction design patterns are the core elements of the two-stage user interface generation and adaptation process as illustrated in figure 2. Therefore, these two pattern categories will be described in more detail in the following sections.

Generic design patterns. Generic design patterns define the global settings of the user interface. These global settings encompass user interface variables such as font size or suitable and preferred interaction mechanisms (e.g. voice control, touch, gesture, etc.) which determine the overall appearance of the individualized user interface across all interaction situations. Generic patterns are closely related to

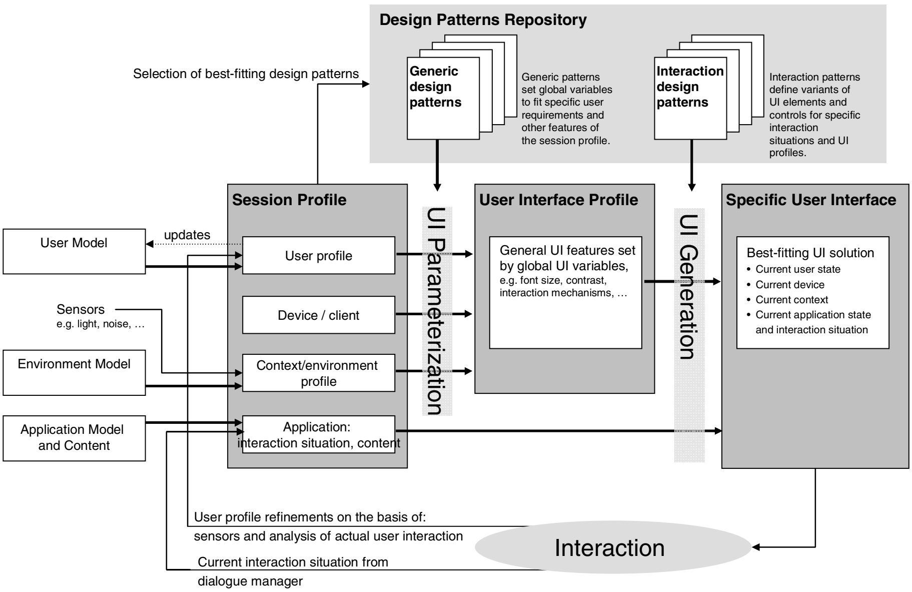


Fig. 2. Generic design patterns and interaction design patterns in the two-stage user interface generation and adaptation process

specific user characteristics as stored in the user profile. They process the current user profile and "translate" user characteristics into user interface features, i.e. global settings. Figure 3 presents an example of a generic pattern which sets the font size in accordance with the current setting in the user profile variable *Visual Acuity and Sensitivity Impairment*.

With a blank user profile (i.e. no information available about the user), the default variants of all generic patterns will be activated. Each time when changes occur to the user profile, all generic patterns will be updated, and the best fitting generic patterns will be activated. This will result in changes to the global settings.

Interaction design patterns. Interaction design patterns provide the interactive elements needed in the current interaction situation as drawn from an abstract application model. For each interaction situation (e.g. main menu, single selection from a list of options) a bundle of interaction design patterns is provided. Depending upon the current values in the global user interface variables (as set by the generic patterns) the most suitable interaction design pattern will be selected from the current bundle in order to cover individual impairments or other constraints.

Interaction design patterns do not have any references to user profile variables. Their input variables are the current interaction situation (problem statement in the pattern description) which is drawn from the abstract application model and relevant user interface settings (context statement in the pattern description). Figure 4 provides an example of an interaction design patterns description.

<ID>	**Font Size - Medium**
Problem	Users with medium vision impairments need moderately increased font sizes in order to be able to read the displayed text in a comfortable manner.
Context	**If** 1 ≤ *Visual Acuity and Sensitivity Impairment* < *2*
Solution	**Then** set *Title Font Size* = 55 points set *Body Text Font Size* = 34 points set *Complementary Text Font Size* = 21 points
Rationale (references)	…
Related patterns	**Substitutes:** "Font Size – Default", "Font Size - Large", "Font Size – X Large" *Title Font Size* **input to** <all interaction patterns with title text elements> *Body Text Font Size* **input to** <all interaction patterns with body text elements> *Complementary Text Font Size* **input to** <all interaction patterns with complementary text elements>

Fig. 3. Example of a generic design pattern: »Font Size – Medium«

Then, from all the selected interaction patterns a complete user interface is generated. At the stage of rendering the user interface, the concrete application content is used as another input source in order to parameterize the user interface

<ID>	**Main Menu – 4x4-Tile Style (Default)**
Problem	Main Menu provides access to all available services/applications of the TV set and might display important status information for some services (e.g. indicate new message for the email service). Note: The displayed items do not have a natural order. Order could be changed by the user or could be subject of adaptations (based on frequency of use).
Context	**If** *Body Text Font Size* ≤ AND *Max. Elements Per Screen* ≥ 15 AND *Cursor Navigation* = on AND *NumericNavigation* = off AND NOT *Display Mode* = text only
Solution	4x4 – define size and positions (grid) of the 16 menu buttons for cursor navigation
Diagram	<Illustrate the design solution in a schematic and concise manner>
Examples	<give examples for situations in which the problem can occur and how it can been solved>
Rationale (references)	<explain the principles or rationale behind the pattern and provide references to literature, standards, etc.>
Related patterns	**Substitutes:** <interaction pattern> **Requires:** Main Menu button **Uses** <*global variable*> **as set by** <generic pattern(s)>

Fig. 4. Example of an interaction design pattern: »Main Menu – 4x4-Tile Style (default)«

widgets, e.g. put specific content options into menu elements and add labels to buttons. Therefore, interaction patterns bridge the gap between the application – namely interaction situations (states) and content elements - and the generic design patterns (generic user interface features). With the input from the application (states and content) and from the generic patterns, suitable interaction patterns are selected and a specific user interface is created.

3.4 Generating and Adapting User Interfaces at Run Time

A User Interface Generator creates a complete user interface from the selected interaction design patterns (i.e. the respective software components). As the main user interaction platform, the MyUI project targets an interactive TV set with the Philips Net TV framework (cf. [21]) which relies on CE HTML and is expected to support more advanced web technologies in the future. For target platform support, the MyUI generator creates web-based user interfaces. Involved technologies include HTML, CSS and AJAX/Java Script. The user interface generator is part of the MyUI middleware which is implemented in PHP (see [17] for a detailed description of the MyUI middleware). As an output of generic and interaction design patterns, a distinct composition of HTML, CSS and Javascript code is being generated.

Relevant changes in the user profile can be recognised and will lead to a reconfiguration of the displayed user interface. Besides explicit user profile configurations in dialogues, possible triggers for changes in the user profile shall be recognized events in the user interaction which indicate problems of use, e.g. repeated time-out events, repeated go-back events, repeated invalid user input, etc. Further input to the user profile is provided by external sensors which capture relevant user behaviour such as facial expressions, head orientation and eye gaze direction, etc. (see [18] for a more detailed description). In order to be able to trigger immediate system reactions to modified user profiles, different client and server sided AJAX components are included into the MyUI adaptation framework.

4 Conclusions

The vision of intelligent user interfaces that adapt to individual needs and preferences is not new. Their realization, however, is a great challenge. The presented concept is an attempt to make use of the widely established notion of interaction design patterns for implementing highly modular and self-adaptive user interfaces that will be able to improve the accessibility of every-day ICT products. The modularity of the approach shall support the adaptation to a maximum heterogeneity of the envisaged target user groups and contexts of use.

We are planning a series of user studies in order to validate the usability and accessibility improvements achieved by the design patterns which have been specified on the basis of research literature and standards. For the design of adaptation mechanisms (the MyUI transition patterns, see above), however, only little knowledge is available. In experimental studies, we will explore approaches that will help to create user trust and acceptance for intelligent interface adaptations. We expect user control to be a major issue for acceptability. Besides automatic user interface adaptation, the MyUI user interfaces can be individualised also manually. This helps users and their relatives, carers or alike to keep control over the user interface appearance. For this purpose, the user profiles as well as the user interface profiles can be viewed, modified and completed manually.

Acknowledgements. The authors acknowledge the help of other partners in the MyUI consortium in this work: Phillips Consumer Lifestyle, Ingenieria Y Soluciones Informaticas Del Sur (ISOIN), CleverCherry.com, Universidad Carlos III de Madrid, The University of Nottingham, Forschungszentrum Informatik, Birmingham City Council, Semmelweis Egyetem, and Ayuntamiento De Getafe.

References

1. Nielsen, J.: Usability Engineering. Academic Press, Cambridge (1993)
2. Harper, R., Rodden, T., Rogers, Y., Sellen, A.: Being Human: Human-Computer Interaction in 2020. Microsoft Research Ltd. (2008)
3. Orwant, J.: Heterogeneous learning in the doppelgänger user modelling system. User Modeling and User-Adapted Interaction 4(2), 107–130 (1994)
4. Krulwich, B.: LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data. AI Magazine 18(2), 37–45 (1997)
5. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems: An Introduction. Cambridge University Press, Cambridge (2010)
6. Weber, G., Specht, M.: User Modeling and Navigation Support in WWW-Based Tutoring Systems. In: Jameson, A., Paris, C., Tasso, C. (eds.) Proceedings of 6th International Conference on User Modeling, pp. 289–300. Springer, Wien (1997)
7. Hartmann, M., Schreiber, D., Mühlhäuser, M.: AUGUR: Providing Context-Aware Interaction Support. In: EICS 2009. Ed. ACM. EICS, pp. 123–131 (2009)
8. Thorrtas, C.G., Krogsoeter, M.: An Adaptive Environment of the User Interface of Excel. In: Intelligent User Interfaces 1993, pp. 123–130 (1993)
9. Krzywicki, A., Wobcke, W., Wong, A.: An Adaptive Calendar Assistant Using Pattern Mining for User Preference Modelling. ACM Press, New York (2010)
10. Ringbauer, B., Peissner, M., Gemou, M.: From "design for all" towards "design for one"–A modular user interface approach. In: Proceedings of the 4th Int'l Conference on Universal Access in Human-Computer Interaction (UAHCI 2007), pp. 517–526. Springer, Berlin (2007)
11. Leuteritz, J.-P., Widlroither, H., Mourouzis, A., Panou, M., Antona, M., Leonidis, A.: Development of Open Platform Based Adaptive HCI Concepts for Elderly Users. In: Stephanidis, C. (ed.) Proceedings of 13th International Conference on Human-Computer Interaction (HCI International 2009), Springer, Berlin (2009)
12. Gajos, K.Z., Weld, D.S.: SUPPLE: Automatically Generating User Interfaces. In: IUI 2004. Ed. ACM. IUI, pp. 93–100 (2004)
13. Gajos, K.Z., Wobbrock, J.O., Weld, D.S.: Improving the Performance of Motor-Impaired Users with Automatically-Generated, Ability-Based Interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008), pp. 1257–1266. Association for Computing Machinery, New York (2008)
14. Brusilovsky, P., Maybury, M.T.: From Adaptive Hypermedia To The Adaptive Web. Communications of the ACM 45(5), 31–33 (2002)

15. Findlater, L., Gajos, K.Z.: Design Space and Evaluation Challenges of Adaptive Graphical User Interfaces. AI Magazine 30(4) (2009)
16. Edlin-White, R., Cobb, S., D'Cruz, M., Floyde, A., Lewthwaite, S., Riedel, J.: Accessibility for older users through adaptive interfaces: opportunities, challenges and achievements (2011) (in preparation; to be published in these proceedings)
17. Gacimartin, C., Hernandez, J.A., Larrabeiti, D.: A middleware architecture for designing TV-based adapted applications for the elderly (2011) (in preparation; to be published in these proceedings)
18. Takacs, B., Simon, L.: Sensing User Needs: Recognition Technologies and User Models for Adaptive User Interfaces (2011) (in preparation; to be published in these proceedings)
19. Dieterich, H., Malinowski, U., Kühne, T., Schneider-Hufschmidt, M.: State of the Art in Adaptive User Interfaces. In: Schneider-Hufschmidt, M., Kühne, T., Malinowski, U. (eds.) Adaptive User Interfaces: Principles and practice, pp. 13–48. North-Holland, Amsterdam (1993)
20. Borchers, J.: A pattern approach to interactive design. John Wiley & Sons Ltd., Chichester (2001)
21. `http://www.nettv.philips.com`

User Interface Representation Using Simple Components

Javier Rodeiro Iglesias and Pedro M. Teixeira-Faria

School of Informatics Engineering, University of Vigo, Spain
jrodeiro@uvigo.es
School of Technology and Management - Polytechnic Institute of Viana do Castelo, Portugal
pfaria@estg.ipvc.pt

Abstract. There are several user interface specification models, each other covering several design steps considered on a user interface representation. Two of those steps are the composition and the presentation. This paper shows a study about the steps that are covered by the most known techniques to represent Graphical User Interfaces, focusing mainly in composition and presentation steps. After this, an example of a game is made using DGAUI representation. This game representation is useful to show the real problems when an application is made from abstraction level to a prototype level. This example is useful too as a possible reference system for testing between different representation systems. The conclusion of this paper is that is possible to make a real application with an abstract representation of user interface, using visual simple components and make a prototype from user interface representation.

Keywords: Abstract User Interface Representation, Visual Appearance, User Interface Dialog, DGAUI.

1 Introduction

When referring about user interface is important to note that we can establish two groups clearly differentiated: the CUI (*Command-Line User Interface*) and GUI (*Graphical User Interface*) [27]. CUI generally display only pieces of text and semi-graphical objects, whereas GUI display interaction objects such as icons, check boxes, list boxes, radio buttons and push buttons. Another issue that is important to address is the relative independence between the *interface visual appearance* concept and the *look and feel* concept. *Look and feel* is a term that comprises interface design aspects, such as: colors, shapes, layout and typefaces which represent the *look*. The aesthetic issue has been considered by several authors [2][13][14][24]. The *feel* is represented by the behavior of dynamic elements such as buttons, boxes and menus. Typically, large software companies create their own style guides or standards. [21] refers that *a standard is a requirement, rule or recommendation based on proven principles and practice. Represents a consensus of a group of officially licensed professionals locally, nationally or internationally*.

The process of designing an interface consists on several steps. Those steps can be represented as models. [6] says that "*A model captures some facets of the problem*

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 278–287, 2011.
© Springer-Verlag Berlin Heidelberg 2011

and translates them into specifications". This idea is reinforced by [12] "*The advantage of using models … it provides designers with abstract constructs, making possible to address user interface design issues at different levels of abstraction*". A previous study has been made and it was possible to verify the existence of different approaches to the interface design process. One approach is presented by [18] who describe eight models, characterized as the *interface design steps* that can be used to represent an interface. The models that represent the design steps are:

– Mental Model (MM)
This model belongs to the user domain. [16] says that software interfaces should be designed to help users build productive mental models of a system. There are common design methods employed to support and influence users mental models, such as: simplicity, familiarity, availability, flexibility, feedback, safety, and affordances.

– Task Model (TM)
One reason for the existence of user interfaces lies in the fact that these are the ones responsible for allowing users to perform their tasks. Several user interface developers recognize the importance that future users and their tasks have, as the most important factors to the success of a user interface [9].

– Application Model (AM)
There is an important relationship between this model and the system implementation. It depends on the physical system (e.g. *desktop GUI*, *mobile device*). According with [18] the Application Model is present in all interface models and acts as distinction element between the user interface and the underlying application.

– Dialog Model (DM)
The dialog structure specification is made with this model. Concerning the tasks performed by users is important to distinguish and correctly specify the application tasks and the interface tasks and how they relate to each other to accomplish their objectives.

– Abstract Presentation Model (APM)
This model main characteristic is the responsibility for the system architecture description through the definition of abstract components which make up the interface and their behavior. Usually, are designated as *CIO (Conceptual Interaction Object)* or *AIO (Abstract Interaction Object)* [23].

– Concrete Presentation Model (CPM)
The concrete graphical visual appearance and interface behavior is represented on this model, which contains details about the implementation of each component. It is dependent of the implementation environment.

– Abstract Composition Model (ACM)
With this model it is possible to define the spatial rules between the interface concrete components (obtained previously from the *Concrete Presentation Model* (CPM)) and

their behavior. Basically, this model represents, on an abstract way, the components topological spatial placing inside the final user interface.

– Concrete Composition Model (CCM)

This model includes the concrete values to place each component on the final interface. Those values are calculated from the abstract composition representation. For example, if a component is centered (has the abstract centered property) it is possible to calculate his actual position from the size of his container component.

Over the years, several methods have been used to specify user interfaces [5]. These include *grammars*, *algebraic specifications*, *task description languages*, *transition diagrams*, *state charts* and *interface representation graphs*. The *DGAUI* system that will be here analyzed is applied to interfaces of GUI type, using *transition diagrams* to represent them. In this approach the transitions represent user inputs and the nodes represent interface states. Computer outputs are represented as either annotations to the state or to the transitions. On next section are presented some techniques used to represent GUIs.

2 Techniques to Represent Graphical User Interfaces

According with the ideal nature of an interactive system, this can and should be represented through a visual formalism: *visual* because it is generated, communicated and understood by humans, and *formal* because it will be manipulated, maintained and analyzed by computers [20]. The term *representation* refers to the visual appearance decided to be used to represent components on the screen (e.g. what looks like a button, a text box). The term *specification* usually refers to a declarative model, with well defined and consistent rules to describe the interface appearance and functionality. It is possible to represent an interface using various representation or specification models presented on the literature. A recent study has been made on interface specification models and XML-Compliant languages. Some of those techniques to represent graphical user interfaces are listed on (Table 1) and analyzed according to the *interface design steps*.

Comparing the *DGAUI* specification with some of the latest specification interfaces languages based on XML (XML-UIDL) we can observe that all the other have an incomplete specification of the *Interface Design Steps*. One possible explanation for this is that each of XML-UIDL has been created with an objective and some evolved to be used on a specific purpose. Because of that, became unnecessary to create a specification which embraced all the design steps.

Usually, the XML-UIDL assumes the previous existence of a set of primitives (traditionally known as *widgets* or *toolkit objects*) which merely allows the presentation manipulation. The (Table 1) shows us that the *DGAUI* representation system is the one which best meets the steps of *composition* and *presentation*, regarding the *interface visual appearance*. Considering that, was decided to deepen the knowledge on the structure and functioning of the *DGAUI* system, which is explained in the following section.

Table 1. Interface Specification Models and XML-Compliant Languages Represented by Design Steps

	MM	TM	AM	DM	APM	ACM	CPM	CCM
Models								
ADC [18]	✓	✓	✓	✓	✓	✓		
ADV [18]	✓	✓	✓	✓	✓			
ALV [18]	✓	✓	✓	✓	✓			
Arch [18]	✓	✓	✓	✓	✓		✓	
CNUCE [18]			✓	✓	✓			
Client-Server [18]	✓	✓	✓	✓	✓			
DGAUI [18]	✓	✓	✓	✓	✓	✓	✓	✓
ICO [18]	✓	✓	✓	✓	✓			
IOG [18]			✓	✓	✓		✓	✓
IRG [18]			✓	✓	✓			
Jacob [18]			✓	✓	✓			
Monolithic [18]	✓	✓	✓	✓	✓			
MVC [18]	✓	✓	✓	✓	✓			
PAC [18]	✓	✓	✓	✓	✓			
PAC-Amodeus [18]	✓	✓	✓	✓	✓			
PUM [18]		✓		✓	✓			
Seeheim [18]	✓	✓	✓	✓	✓			
UDM [18]	✓	✓	✓		✓			
UAN [18]				✓	✓			
York [18]			✓	✓	✓			
XML-Compliant Languages								
DISL [19]		✓	✓	✓	✓		✓	
EMMA [28]				✓				
MariaXML [15]		✓	✓	✓	✓	✓	✓	
SeescoaXML [7][22]				✓	✓		✓	
UIML [1][8][25]	✓	✓	✓	✓	✓		✓	
UsiXML [26]		✓	✓	✓	✓		✓	
XAML [10]							✓	
XForms [29]				✓	✓		✓	
XIML [17]		✓	✓	✓	✓		✓	
XUL [11][3]				✓	✓		✓	

3 User Interface Structure in DGAUI

DGAUI establishes a model for specifying interfaces using XML [30] as a representation technique. To specify user interfaces using XML is considered to be one solution for the standardization and interoperability between applications [4]. XML-UIDL languages have the advantage of being transparent to different interface technologies and provide a uniform resource for heterogeneous communication modes. *DGAUI* has the advantage of owning support tools to each *interface design steps*. Basic\ally, it is a design system that supports interface representation and aims to be considered a milestone in the interface design. The *DGAUI* representation system allows the interface designer:

- To make the *definition of the user interface components*, enabling the creation of composition hierarchies and new components definition;
- To indicate individual components composition to form the *concrete* interface, with which the user interacts;
- To represent the *dialog* between components inside the user interface. It is possible to indicate the events to which each component listen and how the user interface responds to these events (by changing the interface components or making calls to the application).

To create an interface using *DGAUI* is necessary to follow several steps as we can see on (Fig. 1) representation.

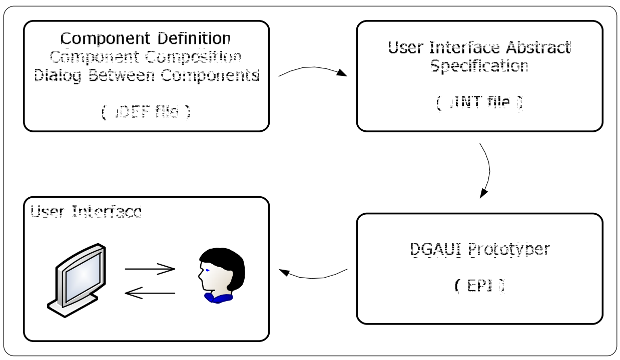


Fig. 1. *DGAUI* interface building process

Initially, the visual components, their composition and the dialogs between them are created. The result is a (.def) file. Then, from that file, the interface designer generates a (.int) file, with the interface states and their transitions. Finally, using an adequate interface prototyper (which reads the (.int) file), the user interface is generated and the user can start interacting with it. This process of creating an interface using the *DGAUI* system is explained below.

3.1 Component Definition (DGAUI-DEF)

The components definition step takes place with the creation of a structure in a XML file, which is divided into three parts:

- Component visual representation

This part is responsible for the components description, which can be described in three ways: *graphic* (based on graphical primitives), *text* (based on text primitives) or *bitmap* (based on pixels). Two of the most important component attributes are: *visible* (indicates the component visibility on the interface) and *active* (indicates if the component responds to the defined events or not). The component start position and size can be fixed or relative to another component (usually the container component). On *DGAUIDE* (an *Integrated Development Environment* designed to support the

DGAUI specification), a *tool palette* is available with tools like: *text*, *rectangle*, *circle*, *ellipse* and *enumeration*. It is possible to access and to change the properties of each component using the appropriate *properties palette*. Depending on the component type, several properties are available, such as: *active*, *color*, *font*, *name*, *visible*. Other operations with components are available like: *elimination*, *copy*, *paste*, *alignment* and *size adjustment*.

Below will be described a user interface prototype of a game on which was used image files (by enumeration). However, the *DGAUI* XML specification contemplates also the use of graphic primitives.

– Component composition

Here is defined the components spatial composition. It is possible to see the abstract position of a component related to other (e.g. *center*, *right align*). The IDE allows the creation of composition relationships between components. This can be done directly by dragging the components or by using the *outline palette*.

– Dialog established between the components

The interaction with the interface is represented on this part, which shows the changes that an event produces over a component itself or in other components. On the IDE it is possible to define which components, previously created responds at which specific events (e.g. *LeftClick*, *RightClick*, *MouseOn*, *Keys*). For each event, the user can specify preconditions and changes (actions) to him. Two event types exist:

- *Component Event* – event related with the actions that a user can do over the visual components (with the mouse and the keyboard);
- *System Event* – no predefined event. Applicable only to the interface in general and not to the components.

The relation between an event over a component or over the interface is called *dialog element* [4].

The resulted interface, obtained from component definition results in a (.def) file, automatically generated by a *DGAUIDE* tool. The (.def) file content is a XML representation of the visual components, component composition and components dialog. Clearly, if we did not have that tool we always could represent the interface simply editing a text file with the XML structure.

3.2 States and Transitions Generation (DGAUI-INT)

The second step of the interface building process is the states and transitions generation. This step takes place after the prior step of component definition. Here, a (.int) file is obtained, including several information about the components visual representation contained in each state, the component composition and the transitions (between states) through by which the interface passes as a result of the dialog set in the (.def) file. It is understood by the state of the interface, that combination of components, at a given time, that is waiting for a user action.

At the same time that the system generates the states, it represents the transitions between them. To represent the states transitions, identifying the user actions that provokes them, it is used a format to indicate a state, an event, the component on which the event applies and the new generated state. Each transition may have an associated set of preconditions for that transition can occur.

Following, an experience building a user interface prototype is described[1].

4 A User Interface Prototype

After an analysis of the *DGAUI* representation system, the next goal was focused on building an interface using simple components. Thus, an idea to create a simple game consisting of sport balls and sport fields have been implemented (Fig. 2).

Fig. 2. A simple game interface

The game rules established were:

- The player must connect the balls with the correct sport fields;
- The player starts by selecting a ball and then choose the corresponding sport field;
- If the user does not make a correct match between the selected ball and the correspondent sport field the game is restarted, "cleaning" all the previous correct connections;
- The game ends when all the connections are made.

For each sport ball, three possible visual states have been created (*normal*, *selected* and *correct*) while in the case of the sport fields two visual states were created (*normal* and *correct*). In terms of component composition, all the interactive elements of the game have been placed inside a container component called *background*.

The prototyper used was EPI (*Interface Proof Evaluator*) [4] and some changes were made to the original code, to enable some game features. Initially, the prototyper receives as input the (.int) file, read it and shows to the user the interface previously defined. The user can interact with the interface and in the example of the game he can use the mouse to click on the balls and the sport fields. The complete functionality of the game can be represented by 20 states (Fig. 3).

[1] Its possible to obtain in http://trevinca.ei.uvigo.es/~jrodeiro/DGAUI/samples the (.int) file correspondent to the game here described and the prototyper (EPI) which shows the game interface.

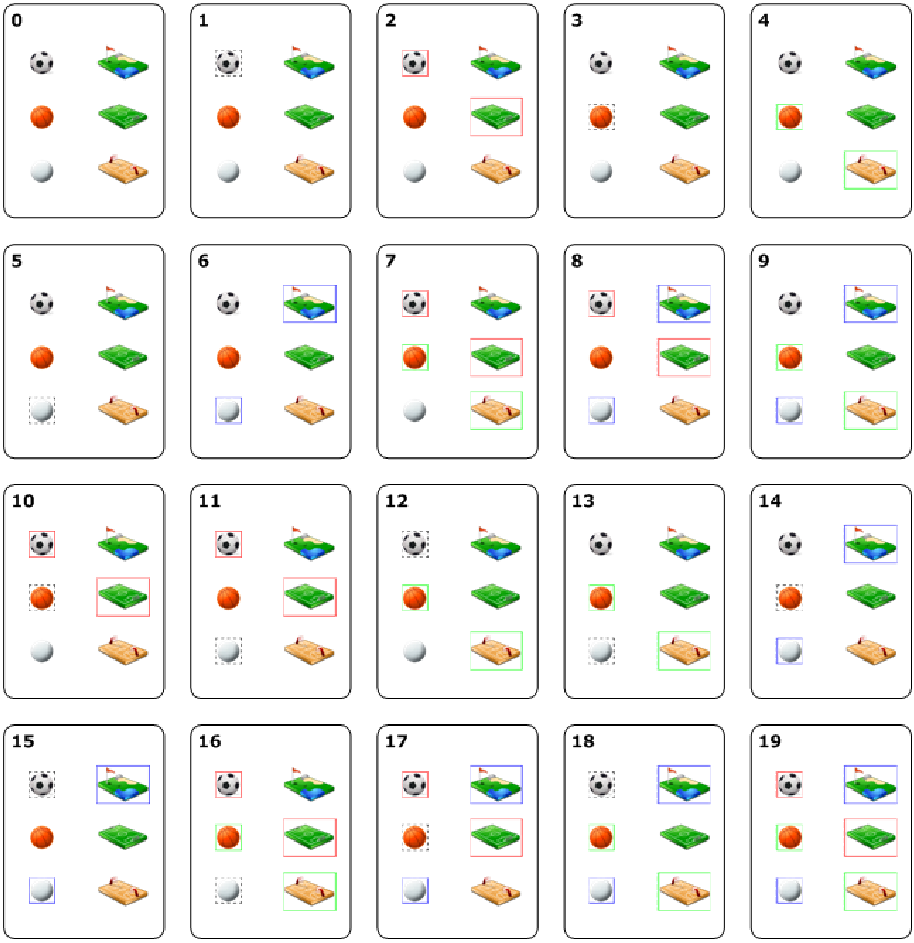


Fig. 3. The 20 interface game states

Using the EPI prototyper is possible to create the *Concrete Presentation Model* correspondent to the interface generated. As stated before, the *DGAUI* system provides concrete component composition, supporting the *Concrete Composition Model.*

5 Conclusions

The idea of this paper was to create a total representation for an interactive visual interface. One of the most important features of an interactive visual interface is the "*visual*". A complete representation of a user interface must include the appearance of it. Most of the representation techniques don't cover this feature. We have used DGAUI as representation to define a simple visual game for kids. It was chosen because, based on a previous study, was concluded that DGAUI was the one which better represents the *Composition* and *Presentation* steps of the interface design process. A concrete interface can be created and putted at the disposal of the user using a proper prototyper (*EPI*). It was possible to verify the implementation of the user interface through a prototype from user interface representation. The example created and here presented is useful as a possible reference for testing different representation systems.

A major conclusion obtained from the work described on this paper is that is possible to make a real application with an abstract representation of a user interface, using visual simple components and make a prototype from representation that could be tested by user. The example here described demonstrated that is possible to build complete interfaces using simple components. For the future, one possible way of study could be increase of the abstraction level on user interface components.

Acknowledgments. This work was supported by:

1. Grant SFRH/PROTEC/49496/2009 of MCTES – Ministério da Ciência, Tecnologia e Ensino Superior (Portugal)
2. Project TIN2009-14103-C03-03 of Ministerio de Ciencia e Innovación (Spain)
3. Project 10DPI305002PR of Xunta de Galicia (Spain)

References

1. Abrams, M., Phanouriou, C.: An XML Language for Building Device-Independent User Interfaces, Philadelphia (1999)
2. Bodart, F., Hennebert, A., Leheureux, J., Vanderdonckt, J.: Towards a Dynamic Strategy for Computer-Aided Visual Placement. ACM, Bari (1994)
3. Bojanic, P.: The Joy of XUL (2006), `http://developer.mozilla.org/en/docs/TheJoyofXUL`
4. Carnero, S.: Sistematización de la validación de Interacción del usuario sobre la visualización en Interfaces de Usuario usando Especificación Abstracta. PhD Thesis, Universidad de Vigo (2007)
5. Carr, D.: Specification of Interface Interaction Objects. In: CHI 1994 – ACM Conference on Human Factors in Computer Systems, pp. 372–378 (1994)
6. Diaper, D., Stanton, N.: The Handbook of Task Analysis for Human Computer Interaction. Lawrence Erlbaum Associates, Mahwah (2004)
7. Guerrero-Garcia, J., Gonzlez-Calleros, J., Vanderdonckt, J., Muoz-Arteaga, J.: A Theoretical Survey of User Interface Description Languages: Preliminary Results. IEEE Computer Society, Los Alamitos (2009)
8. LiquidApps (2009), `http://www.liquidappsworld.com/`
9. Meixner, G., Thiels, N.: Tool Support for Task Analysis (2008)
10. Microsoft Corp.: XAML (2010), `http://msdn.microsoft.com/enus/library/ms752059.aspx`
11. Mozilla Developer Center: XUL (2010), `https://developer.mozilla.org/en/XUL`
12. Navarre et al.: ICOs: a Model-Based User Interface Description Technique dedicated to Interactive Systems Addressing Usability, Reliability and Scalability. In: ACM ToCHI (2009)

13. Ngo, D., Byrne, J.: Another look at a model for evaluating Interfaces Aesthetics. Int. J. Appl. Math. Computer Science 11(2), 515–535 (2001)
14. Ngo, D., Teo, L., Byrne, J.: Evaluating Interface Esthetics. Journal Knowledge and Information Systems 4(1) (2002)
15. Paterno', F., Santoro, C., Spano, L.D.: MARIA: A Universal, Declarative, Multiple Abstraction-Level Language for Service-Oriented Applications in Ubiquitous Environments. ACM Transactions on Computer-Human Interaction 16 (2009)
16. Preece, J.: Human-Computer Interaction. Addison-Wesley, Reading (1994)
17. Puerta, A., Einstein, J.: XIML: A Universal Language for User Interfaces. Red-Whale Software (2001)
18. Rodeiro, J.: Representácion y Análisis de la Componente Visual de la Interfaz de Usuario. PhD Thesis, Universidad de Vigo (2001)
19. Schaefer, R., Bleul, S., Müller, W.: Dialog modeling for multiple devices and multiple interaction modalities. In: Coninx, K., Luyten, K., Schneider, K.A. (eds.) TAMODIA 2006. LNCS, vol. 4385, pp. 39–53. Springer, Heidelberg (2007)
20. Schlungbaum, E., Elwert, T.: Dialogue Graphs - A Formal and Visual Specification Technique for Dialogue Modelling. In: BCS-FACS Workshop on Formal Aspects of the Human Computer Interface, Sheffield Hallam University. Springer, Heidelberg (1996)
21. Smith, J.: ISO and ANSI ergonomic standards for computer products. A guide to implementation and compliance. Prentice-Hall, Englewood Cliffs (1996)
22. Souchon, N., Vanderdonckt, J.: A Review of XML-compliant User Interface Description Languages, pp. 377–391. Springer, Heidelberg (2003)
23. Szekely, P.: Retrospective and Challenges for Model-Based Interface Development, pp. 1–27. Namur University Press (1996)
24. Tractinsky, N.: Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In: CHI 1997, Atlanta, USA (1997)
25. UIML3.1 Draft specification (2009), `http://www.uiml.org/specs/uiml3/DraftSpec.htm`
26. USIXML (2009), `http://www.usixml.org/`
27. Vanderdonckt, J., Gillo, X.: Visual Techniques for Traditional and Multimedia Layouts. ACM, Bari (1994)
28. W3C Recommendation: EMMA (2009)
29. W3C Recommendation: XForms 1.1 (2009)
30. W3C Recommendation: XML, XML 1.0, 5th edn. (2008)

Model-Based Ubiquitous Interaction Concepts and Contexts in Public Systems

Thomas Schlegel and Christine Keller

Technische Universität Dresden, Germany
{Thomas.Schlegel,Christine.Keller}@tu-dresden.de

Abstract. Ubiquitous systems and interaction concepts are increasingly finding their way into public systems like shopping malls, airports, public transport or information kiosks. At the same time, these user interfaces also undergo significant changes. Technologies like multi-touch systems or voice-based interaction are now available to the general public and widely used. In ubiquitous systems, these modalities are often combined, sometimes even dynamically at runtime. This leads to new challenges for the conceptualization and development of ubiquitous user interfaces in public systems, especially where this implies adaptive behavior. We present contexts that possibly influence the interaction with such public systems and describe ways of modeling this interaction integrating context-adaptivity already in the interaction models of public systems. Taking into account the context of the public system and its users, we extend the concept of Interaction-Cases to contain model aspects for different interaction contexts in public systems.

Keywords: Interaction Workflow Modeling, Context, Context Models, Context Adaptivity, Public Systems, Use Cases, Interaction Cases, Interaction Models.

1 Introduction

Public systems represent a field of application for ubiquitous techniques that is growing recently. Ubiquitous environments are integrated into public information technology and socio-technical public systems like shopping malls, airports or public transport[1,2]. Some of them have particular target groups, like tourists or handicapped people [3,4].Others are designed to serve the general public[5]. New interaction technologies like multi-touch interfaces or gesture recognition facilitate the construction of really pervasive computing environments for public systems.

In public places, computing environments must be as unobtrusive as possible. At the same time, their usability and accessibility must be very high and designed to support a variety of strongly different users without training. Combining multiple interaction techniques in ubiquitous environments for public systems aims at supporting both, pervasive and easily accessible computing that is optimized to the situation of the user without the availability of concepts like personalization, profiles etc. found in "classical" interactive systems like business software, home automation or control environments. Ubiquitous systems offer new opportunities for public

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 288–298, 2011.
© Springer-Verlag Berlin Heidelberg 2011

settings, which will make them find their way into public systems, but also bring new challenges with them, regarding their development and modeling.

Ubiquitous systems support many different interfaces and often need to be context-aware to optimize their interactive behavior regarding context. They are characterized by integrating several means of interaction and being highly adaptive. Here, a model-based approach can lead to efficient development of ubiquitous systems and support adaptive design of user interfaces in ubiquitous systems [6]. However, model-based design methods like Model-Driven Architecture (MDA) are often not flexible enough to meet the needs of designing adaptive, context-aware ubiquitous systems[7]. They also focus on design-time generative approaches, which are not applicable to non-monolithic, modular systems with changes on runtime. Therefore, a modeling-technique for interactive public systems is needed, which supports context-aware and highly adaptive interactive ubiquitous systems, in order to facilitate the creation of ubiquitous public systems that combine context-adaption as well as multiple and multimodal interaction techniques.

In traditional software engineering, user and interaction models are specified at design time, using persona, textual descriptions or just the mental models that user interface designers and developers have about the future users of the system[8,9]. Based on these user models and the identified roles and actors, the whole system is then designed[10,11]. In ubiquitous public systems, however, the users can often not be associated with a specific user group. In public, all kinds of people are around and can turn into users of the public system, often with different or unspecific needs not covered with a dedicated task or system, making the user anonymous and unpredictable at design-time. However, the context of the public system and its user interface is known, can be specified and observed at runtime. Based on the context the public system observes, such as location, time or input, the system can adapt, for example, its interaction modalities. Therefore, especially in public systems, the modeling of an interactive and ubiquitous system has to focus on system context.

The field of context-aware systems is very broad, being researched for several years now. Especially the growth of interest in ubiquitous systems has pushed the development of context-aware environments[12,13]. One of the first context-aware systems was the *active badge location system*, developed by Want et al. in 1992[14]. It observed the user's location and redirected telephone calls concerning this location. The early context-aware systems mainly considered location as the context of the user, like many tourist guides, for example[3,15]. Later on, other contexts were considered as well, which led to different approaches on modeling context. Dey et al. for example, identified several popular categories of context such as "location, identity, activity and time"[13]. The classification of context that is necessary for building context-aware systems led to the development of several ontologies for context, for example by Chen et al. and also by Moore et al. [16,17].Many context-aware ubiquitous systems emerged and this development led to the design of several frameworks that facilitate the construction of context-aware systems[18,19].

The specifics of public systems were considered in some ubiquitous systems that were designed for public environments, for example, systems that support handicapped people in settings like public transport[4]. Other systems, like the GUIDE project, focused on tourists, or on students on a campus, like the e-campus project [2,3]. Some other projects involving public transport settings used ubiquitous and context-aware technologies [15,20]. However, to our knowledge there is little research about the properties and contexts of public systems that does not focus on specific settings or user groups, like tourists, students, public transport etc. We therefore define a general public system as a system that performs in public spaces and does not target specific user groups but is available to all people, i.e. the public.

Concerning modeling techniques for interactive systems, there is extensive work on MDA and model-driven interface design[6,7]. As noted above, some extensions of UML have led to modeling languages for web applications or web services, some of them context-aware[21,22]. Most of this work, however, does not consider context-awareness and adaptivity in ubiquitous public systems.

In this paper we therefore present an overview on possible contexts in public systems and a basic taxonomy of these contexts. We then introduce a model-based approach on designing context-adaptive interaction in ubiquitous public systems.

2 Dimensions of Context in Public Systems

Since in public systems, personalization often is impossible, they strongly depend on context in order to adapt to user's needs and the surroundings. Models of ubiquitous public systems should be adaptive regarding these different contexts. As a basis for adaptive modeling techniques, the possible contexts of public systems need to be analyzed. We developed a classification of contexts of public systems for this purpose, which is shown in figure 1. The context classification is still work in progress and should not be too fine-grained for serving as a basis for different systems and context ontologies. It shows our basic approach to modeling context and points out the specific properties of public systems that must be taken into account while modeling interactive ubiquitous systems for public settings. Ideally, context models are defined in a domain-specific way for the specific public system, but rely on a common basic ontology that allows for matching and integrating the context ontologies on the top level of abstraction. The examples of context-adaptive interaction models for ubiquitous public systems, whichwe describe in section 5, show that already a coarse-grained and incomplete context classification can be applied successfully to our modeling technique.

Most ubiquitous systems have a kind of **InteractionContext** (**I**), consisting of the system's *InputContext*(*In*), representing the possibilities of input, i.e. of entering, selecting or editing information. The system also has a *ProcessingContext*(*Proc*) that reflects the processing of input on part of the system, including capability models and sensor fusion. Analogously, there also exists an *OutputContext* (*Out*), which relates to different output modalities. Especially in public areas, the context of interaction by the user has to be considered as well, in particular as there are aspects that are known to the system without specific knowledge about the user. A user not only perceives and acts while interacting with a public system; he also processes what he has perceived. This leads to the *Perceptive*, *Cognitive* and *ActingContext*(*Per, Cog, Act*)of the user as important contextual aspects. In public, the user can be distracted by noise, too much light (e.g. glare) as well as bad lighting, for example. These conditions influence his *Perceptive* as well as his *CognitiveContext*, like being in a hurry or looking for somebody else, which will reduce or divert his attention and therefore will reduce his available cognitive and perceptive capacity.

Of course, the **Spatial Context** (**Sp**) must be considered, too. This concerns the location of the user as well as the location of the system. Large public displays as well as the mobile devices of users may be a component of a public system. The location of the system therefore can be *fixed (Fix)* or *moving (Mov)*, but can also influence size and visibility of an available display. We furthermore identified the **Temporal Context** (**Temp**) with its sub-contexts *AbsoluteTime* (*Abs*) and *RelativeTime* (*Rel*). Relative time will occur, for example, where a distance separates the user from a destination, leading to a relative time needed to get there depending on transportation, or where interaction occurs relative to an event like the late arrival of a train.

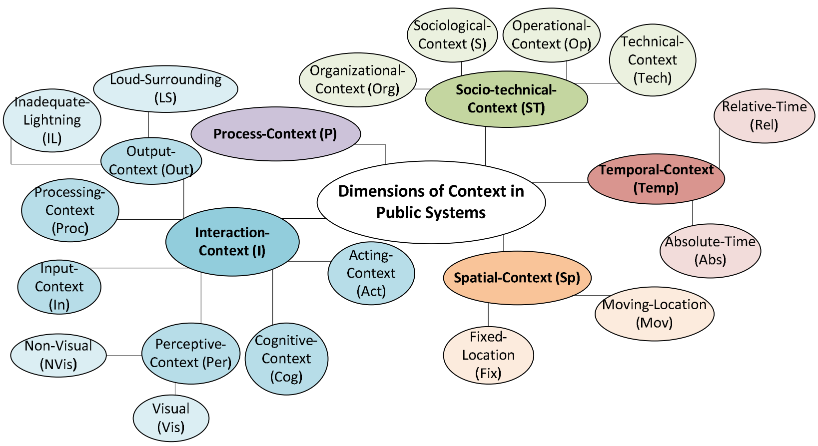


Fig. 1. Dimensions of Context in Public Systems

A context specific to and of high interest in public systems is the **Socio-TechnicalContext** (**ST**). In public systems, there is a *SociologicalContext* (*S*), considering the common use of resources by many people. There are social rules in public spaces that are followed by most of the people there. These rules can affect the usage of public ubiquitous systems. Concerning, for example, the interaction using a big public display, the aspect of privacy has to be considered. A user should not be required to enter or request personal data in such a way that might disclose it to others[23].

The *TechnicalContext* (*Tech*) is also a sub-context of the Socio-Technical Context and describes the technical abilities of the system. In public transport this could be the question whether a subsystem has fast and reliable access to an up-to-date timetable data source, including delays or failures, or only to the regular timetable.

Another sub-context of the Socio-Technical Context of public systems is the *OperationalContext* (*Op*) of a system. This covers everything that goes on "behind the scenes" and includes, in public transport for example, activities in the railway control center. This context affects public systems, because many of them depend on certain operational procedures and exhibit only a small part of the full socio-technical system to the public.

The *OrganizationalContext* (*Org*) models the conditions and contexts of organizations involved. In public spaces, these could be the operating company of public transport as well as organizations running shops at airports or railway stations. For services, the Organizational and the whole **ProcessContext (P)** are of high importance. Process Context in public systems is defined upon user and system actions and their structural and logical interdependencies (causality etc.). Many actionsexhibit dependencies or will trigger other actions necessary to accomplish a complete task with the desired result. Especially regarding service-oriented architectures, the Process Context is of high value for service quality and many other aspects like security and safety.

This also shows that many of the context types, like Organizational Context and Technical Context are interconnected or may even be integrated to form, in this example, the Socio-Technical Context. Therefore, one must consider the modeling of contexts as well as their integration or even fusion.

3 Interaction-Cases

When software systems with interactive components are designed, often stakeholders from different disciplines participate in the process. The design step therefore requires techniques to be easily understandable as well as highly applicable in different phases and aspects from informal requirements to formal processes, in order to be used by technical as well as non-technical stakeholders. Paper-based prototypes, scenarios and textual descriptions of a system's behavior are non-technical and easily understandable approaches for early design phases[10,24]. However, these artifacts lack formalization and therefore cannot be linked to artifacts of later development stages like code. Often the early stage design artifacts become outdated and inconsistent due to changes in later artifacts, which cannot be tracked back to the descriptions in early specification artifacts.

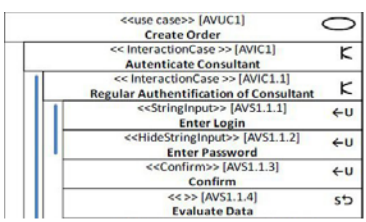


Fig. 2. Part of the specification of an authentication process: Interaction-Case "Authenticate Consultant" within the Use Case "Create Order"

The Unified Modeling Language (UML) provides a semi-formal way of modeling and specifying software in several development stages, from informal Use Case definitions to class diagrams and generated code stubs. Standard UML Use Cases provide a non-technical means of specifying the behavior of a software system. Since they are based on textual descriptions, they are easy to understand also for non-technical participants of the design process. The Use Cases can then be used to create more formal specifications, like class diagrams and could serve as a link to a formalized system specification.

Interactive components, however, are still characterized as plain textual descriptions within Use Cases. In earlier work we therefore proposed **Interaction-Cases** as a technique based on Use Cases, designed to support an easily applicable and at the same time formally strong means of specification of interaction procedures[25]. Interaction-Cases were created to support incremental and iterative development and to enable developers and designers to start modeling on paper, being able to formalize and refine the Interaction-Cases in later stages of development and link them to other artifacts of the development process.

An Interaction-Case consists of sequences of Interaction-Cases and/or (atomic) Interaction-Steps. By nesting Interaction-Cases, a sequence of interactions can be modeled coarse-grained in early design stages and can then be refined into more detailed Interaction-Cases or, vice versa, a sequence of atomic Interaction-Steps can be aggregated later on. Interaction-Cases and Interaction-Steps have an identity and can therefore be referenced and reused by other Interaction-Cases using this identifier. Interaction-Case components can inherit from other components, thus introducing object-oriented type semantics and enabling powerful re-use. The parent class of an Interaction-Case can be noted like an UML stereotype (<<type>>), as shown in figure 2.

Applying the object oriented paradigm further, abstract Meta Components are introduced, identified by a "?" in front of their name. The concept of Meta Components is similar to the concept of interfaces or abstract classes in object oriented programming: A Meta Component specifies certain characteristics of an Interaction-Case, but must later be realized by inheritance to form a concrete one.

The steps of an Interaction Case can be modeled to be executed sequentially, in parallel or without a predefinition of the execution order at design time.

It is possible to direct the flow of Interaction-Cases in branches. There are three possible types of branches. Conditional branching initiated by the user (?DECIDE), initiated by the system (?ConditionalBranch)and unconditional branching by the system (?GOTO). Conditions for conditional branches are followed by a question mark. In parallel execution it is possible to execute all branches, m out of n and to model options for the user to decide (at least one, not all or one out of n).

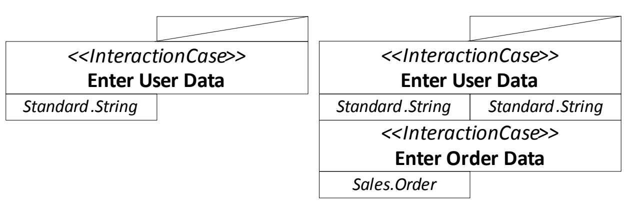


Fig. 3. Interaction-Cases with Input- and Output-Container

In addition to the concepts described in [25], we extended this basic definition of Interaction-Cases by adding an Input and a n Output Container describing the data type of the component's input and output (ObjectType). The Containers are visualized by boxes on top of and below an Interaction-Case, as shown in figure 3. The top box describes the input consumed by a component, e.g. the data that is presented to the user in the interactive step, while the box on the bottom of the Interaction-Case or Interaction-Step describes its output, e.g. the information gathered from the user or read from a sensor or database. The ObjectTypes of input and output of a component are inherited from its parent class. If no input or output is defined, the respective box is crossed out (figure 3). The drawing also shows that the input ObjectType of a successor has to match the output ObjectType of its predecessor. The visualization of input and output-Containers easily ensures consistency of information objects even in early sketching phases.

Also, an extension to the Microsoft Visual Studio 2010 Editor for UML Use Case diagrams was developed, which facilitates the integrated modeling of Interaction-Cases and Interaction-Steps within the development environment. Figure 4 shows a screenshot of the editor extension that was developed.

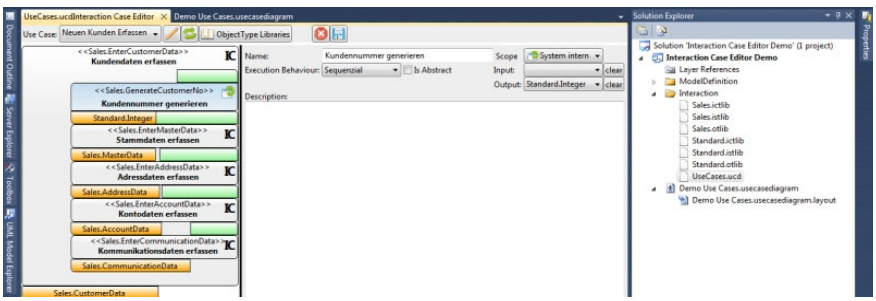


Fig. 4. Screenshot of the extension of the Visual Studio Use Case Editor

Types of Interaction-Steps or Interaction-Cases are defined by inheritance from a parent Interaction-Step or Interaction-Case. Type definitions for Interaction-Step types and Interaction-Case types can be stored in libraries and imported into other projects as well, to facilitate reuse. Based on the present possibilities for modeling Interaction-Case types, Interaction-Step types and ObjectTypes, the generation of code stubs becomes possible, which would further increase the integration of Interaction-Cases into the development of interactive systems.

4 Context-Adaptive Interaction-Cases

In order to support ubiquitous public systems with context-based modeling, we extended the Interaction-Case modeling technique so that Interaction-Cases can be modified by certain Context Types. In early development stages it is possible to denote Interaction-Cases without defining all Interaction-Steps, which supports incremental modeling. Already in this phase, Interaction-Cases can be labeled with

Context Modifiers, to indicate that the refinement of this Interaction-Case depends on certain types of context. In the visual representation of an Interaction-Case, a Context Modifier is indicated by the abbreviation of the Context Type, written in a semi-circle shape that is placed on the left side of the Interaction-Case drawing, as shown in figure 5 on the left.

Alternative Interaction-Cases can now be modeled for different instantiations of the denoted Context Type. This allows to model different context-dependent forms of system behavior. In order to allow Interaction-Cases to be substituted by their context-modified equivalent, these equivalents should be specializations of a common Interaction-Case super-type. Of an Interaction-Case that is modified by the Perception Context (**Per**), alternative Interaction-Cases can be modeled, for example for the Context Types Visual (**Vis**) and Non-Visual (**NVis**). This allows, for instance the modeling of a ubiquitous public system that switches to non-visual (e.g. speech-based) interaction, in case the Perception Context of the user is Non-Visual. This can occur if the user is blind, visually impaired or maybe running towards a train, looking for the right number.

Context-adaptive systems should of course not only adapt to one type of context at a time. Therefore, it is possible to model Interaction-Cases that have multiple Context Modifiers. In that case, the abbreviations of the different contexts that influence the interaction component are noted within the semi-circle shape. For different combinations of Context Types, different modified Interaction-Cases can be modeled. However, since many different combinations of Context Types are possible if there is more than one Context Modifier, not all combinations must be modeled as separate Interaction-Cases. The system or the developer can choose the most specialized context-modified Interaction-Case available that fits the context currently observed.

An example would be an Interaction-Case that has a Context Modifier based on Perception Context (**Per**) and also a Context Modifier based on Output Context (**Out**). It is now possible to model several kinds of Interaction-Cases. One Interaction-Case could be specified for Non-Visual (**NVis**) as sub-context of Perception Context and Loud-Surrounding (**LS**) as sub-context of Output Context. If a blind or visually impaired person would have to use a public system in a loud surrounding, this system then could adapt by not only choosing speech-based interaction, but also by increasing the volume of its speech output as well as the sensitivity and noise reduction of the microphone. Another kind of Interaction-Case could be modeled for Visual (**Vis**) as sub-context of Perception Context and for Inadequate-Lighting (**IL**) as sub-context for Output Context. Instead of switching to speech-based interaction, this Interaction-Case could be realized using enhanced screen brightness.

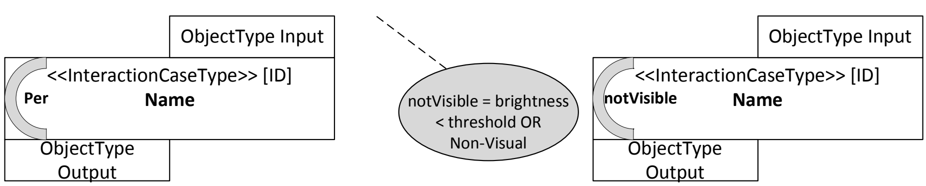


Fig. 5. Context-adaptive Interaction-Case and usage of rule-based sub-context

Exploring the combination of sub-contexts further led us to introduce rule-based sub-contexts. These contexts refine Context Types like Inadequate Lighting or Loud Surroundings, by introducing the parameters that concern these Context Types. It is then possible to define a sub-context of Inadequate Lightning that states “brightness < threshold”. Using such a rule-based sub-context, the abovementioned scenario can be modeled more precisely. The second Interaction-Case can be specialized for “brightness >= threshold”, when enhancing the screen brightness is still effective. An additional Interaction-Case using “brightness < threshold” could then switch the interaction to speech-based.

The combination of contexts so far included combinations using AND logic of multiple inheritance. With the use of rule-based contexts, context combinations can be extended to rule-based sub-contexts like “context1 OR context2”, using logical operators different from AND. The logical operators permitted here are OR/NOT/XOR, further logical operators would complicate the modeling too much. In our modeling technique, rules like that are modeled as separate contexts and also used as separate contexts, too. However, they still are modeled as specializations of the basic context types seamlessly integrating them in the taxonomy of context available.

This way, the aforementioned example Interaction-Cases can be expanded, using the rule-based context not Visible = “((brightness < threshold) OR Non-Visual)”. Both context conditions then lead to speech-based interaction. This example is also shown in figure 5.

5 Discussion and Future Work

In this paper we have presented context-adaptive Interaction-Cases as a modeling technique that supports the modeling of interactive context-aware and adaptive systems. Using our classification of context in public systems, this modeling technique can be used for supporting the development of ubiquitous public systems, as they emerge in many different public areas, like airports or shopping malls. The modeling technique of Interaction-Cases can already be used in the very early stages of system design and can be applied only using pen and paper. Its design, however, supports the transfer of paper-based models to more formal descriptions. In order to demonstrate this, an extension for the Microsoft Visual Studio 2010 Editor for UML Use Case diagrams, that enables the modeling of Interaction-Cases and Use-Cases in the development environment, has been developed. The modeling technique also supports the iterative development of the interaction model by providing inheritance and specialization of types, incorporating the powerful object oriented paradigm.

We are planning on refining our context classification of public systems in order to better support the development of ubiquitous and context-aware public systems. Our Microsoft Visual Studio 2010 Editor extension is planned to be extended to include our context-adaptive enhancement of Interaction-Cases. We are also planning on studying earlier development phases and to integrate even less formal conceptualization and prototyping techniques into a model-based design process that allows the easy but also powerful construction of ubiquitous interactive systems.

Acknowledgement. Part of this work has been executed under the project IP-KOM-ÖV funded by the German Federal Ministry of Economics and Technology (BMWi).We wish to thank Tobias Grass for the contributions made by his thesis and the colleagues from the Institute for Visualization and interactive Systems (VIS).

References

1. Peterson, M.: Pervasive and Ubiquitous Public Map Displays (2004)
2. Storz, O., Friday, A., Davies, N., Finney, J., Sas, C., Sheridan, J.: Public Ubiquitous Computing Systems: Lessons from the e-Campus Display Deployments. IEEE Pervasive Computing 5(3), 40–47 (2006)
3. Cheverst, K., Davies, N., Mitchell, K., Friday, A.: Experiences of developing and deploying a context-aware tourist guide: the GUIDE project, pp. 20–31 (2000)
4. Klante, P., Krösche, J., Boll, S.: AccesSights – A Multimodal Location-Aware Mobile Tourist Information System. In: Miesenberger, K., Klaus, J., Zagler, W., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 287–294. Springer, Heidelberg (2004)
5. Bertolotto, M., O'Hare, G., Strahan, R., Brophy, A., Martin, A., McLoughlin, E.: Bus Catcher: a Context Sensitive Prototype System for Public Transportation Users, pp. 64–72 (2002)
6. Pérez-Medina, J.-L., Dupuy-Chessa, S., Front, A.: A Survey of Model Driven Engineering Tools for User Interface Design. In: Winckler, M., Johnson, H., Palanque, P. (eds.) TAMODIA 2007. LNCS, vol. 4849, pp. 84–97. Springer, Heidelberg (2007)
7. Kruchten, P., Obbink, H., Stafford, J.: The Past, Present, and Future for Software Architecture. IEEE Software 23(2), 22–30 (2006)
8. Pruitt, J., Grudin, J.: Personas: Practice and Theory, pp.1–15 (2003)
9. Wasserman, A.: User Software Engineering and the design of interactive systems, pp. 387–393 (1981)
10. Aoyama, M.: Persona-Scenario-Goal Methodology for User-Centered Requirements Engineering. In: IEEE International Conference on Requirements Engineering, vol. 0, pp. 185–194 (2007)
11. Kazman, R., Abowd, G., Bass, L., Clements, P.: Scenario-based analysis of software architecture. IEEE Software 13(6), 47–55 (1996)
12. Prekop, P., Burnett, M.: Activities, context and ubiquitous computing. Computer Communications 26(11), 1168–1176 (2003)
13. Dey, A., Abowd, G.: Towards a better understanding of context and context-awareness (2000)
14. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. ACM Transactions on Information Systems 10(1), 91–102 (1992)
15. Hristova, N.: Ad-Me: A ContextSensitive Advertising System, pp. 10–12 (2001)
16. Chen, H., Finin, T., Joshi, A.: An ontology for context-aware pervasive computing environments. Knowl. Eng. Rev. 18(3), 197–207 (2003)
17. Wang, X., Zhang, D., Gu, T., Pung, H.: Ontology Based Context Modeling and Reasoning using OWL, p.18 (2004)
18. Salber, D., Dey, A., Abowd, G.: The context toolkit: aiding the development of context-enabled applications, pp. 434–441 (1999)
19. Biegel, G., Cahill, V.: A Framework for Developing Mobile, Context-aware Applications, pp. 361 (2004)

20. Banâtre, M., Couderc, P., Pauty, J., Becus, M.: Ubibus: Ubiquitous Computing to Help Blind People in Public Transport. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 310–314. Springer, Heidelberg (2004)
21. Ceri, S., Daniel, F., Matera, M., Facca, F.: Model-driven development of context-aware Web applications. ACM Trans. Internet Technol. 7 (2007)
22. Sheng, Q., Benatallah, B.: ContextUML: A UML-Based Modeling Language for Model-Driven Development of Context-Aware Web Services Development, pp. 206–212 (2005)
23. Pernack, R.: Public space and transport: a sociotheoretical approach. Wissenschaftszentrum Berlin für Sozialforschung gGmbH (2005)
24. Barbosa, S.D.J., Paula, M.G.: Interaction Modelling as a Binding Thread in the Software Development Process (2003)
25. Schlegel, T., Raschke, M.: Interaction-Cases: Model-Based Description of Complex Interactions in Use Cases (2010)

Towards Pattern-Driven Engineering of Run-Time Adaptive User Interfaces for Smart Production Environments

Marc Seissler[1], Kai Breiner[2], and Gerrit Meixner[3]

[1] University of Kaiserslautern, Center for Human-Machine-Interaction,
Gottlieb-Daimler Str. 42, 67663 Kaiserslautern, Germany
Marc.Seissler@mv.uni-kl.de
[2] University of Kaiserslautern, Software Engineering Research Group,
67663 Kaiserslautern, Germany
Breiner@cs.uni-kl.de
[3] German Research Center for Artificial Intelligence (DFKI),
Trippstadter Str. 122, 67663 Kaiserslautern, Germany
Gerrit.Meixner@dfki.de

Abstract. Model-based development of run-time adaptive user interfaces still poses a complex problem since several models have to be specified and interlinked by the developers. In this paper we present a first concept of a pattern driven development methodology that addresses the reuse of user interface aspects. An XML-based pattern notation is introduced that supports the specification of patterns for different UI core models (room-based Use Model, Dialog Model and Presentation). The pattern notation further separates the pattern interface description from the used model fragments via a pattern interface implementation. This enables the light-weight integration of a pattern application process without touching the used core models' specifications.

Keywords: Model-Based User Interface Development, HCI Pattern, Engineering Run-Time Adaptive User Interfaces, Usability Engineering.

1 Introduction

With the increasing number of mobile interaction devices (like smartphones and Tablet PCs) and wireless communication technologies new forms of human-machine interaction have evolved. Run-time adaptive Universal Control Devices (UCD) such as the SmartMote [1, 2] can help to reduce operating errors in future ambient-intelligent industrial production environments – like the *SmartFactory*^{KL} [3] – by facilitating a single homogeneous user interface for heterogeneous field devices.

While the users benefit from these new technologies they imply a new form of complexity for the developers. To address this problem, new development methodologies, like the model-based user interfaces development (MBUID) [4] have been introduced, which promise to provide a remedy. Using models allows the abstraction from implementation details and a better separation of concerns. This results in a better maintainability and further enables to use more enhanced concepts of reuse.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 299–308, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Especially in the domain of MBUID patterns have been used to further reduce the development complexity – see e.g. [5, 6] – and for teaching design principles [7] by reusing user interface artifacts. Since patterns bridge many levels of abstraction [8] they can perfectly be used as flexible "elementary building blocks" [9] throughout a model-based development process. The notion of variety of these building blocks allows their problem specific instantiation and application on the user interface core models.

But although patterns have gained a lot of interest in the MBUID domain, there still is no common understanding about the activities involved in creating and integrating HCI patterns in current pattern tools and how to use these tools effectively [10].

In this article we introduce a concept for the integration of HCI patterns in our model-based development approach for the systematic engineering of run-time adaptive user interfaces. The rest of the paper is structured as follows.

In the next section we present related pattern-driven development approaches which have been introduced in the domain of MBUID. After that we introduce the initial version of our pattern framework, starting with the user interface core models of our MBUID approach. Based on that, we focus on the pattern notation which is used for the machine-processable formalization of the user interface. While this paper presents a work in progress we will highlight the pattern instantiation phase via a first example user interface. After that, we will discuss our lessons learned and outline the upcoming steps which will be addressed in the future work.

2 Related Work

Originally, patterns were introduced by Alexander [11] for the systematic development of urban architecture. After that the concept has been adopted in various domains like Software Engineering [12] and design of interactive systems [13, 14].

Over the last decade "Generative Patterns" [15] also have been used in model-based user interface development approaches, where they served as reusable building blocks. Compared to non-formalized patterns, generative patterns in MBUID facilitate a higher degree of formalization which might limit the range of application but allows a better machine-processing and automatic application of the pattern. Because of this limitation, these patterns are also considered as *Pattern Instance Components* (PIC) [16] describing one rather concrete instance of a generic HCI pattern. Therefore, PICs might be interpreted as attributed templates that can be instantiated in different phases within the MBUID process. Different levels of abstraction – e.g. as proposed in the CAMELEON Reference Framework [4] – have to be covered by a pattern based approach to support the systematic user-centered UI development. Pattern application frameworks and tools have been introduced which (partially) cover the four abstraction levels of this framework.

Many approaches have been focusing on identifying and integrating patterns in the early stages of the design process. In [17] Paternò introduced a limited set of task patterns, formalized in the *ConcurrTaskTree* (CTT) Notation. Based on these first task patterns, languages like the XML-based *Task Pattern Markup Language* (TPML) [18] have been introduced that formalized the concept in a machine-processable format. Based on this language, tools like the *Task-Pattern Wizard* [6],[18], have been introduced that support developers in applying these patterns in a development process.

Later pattern approaches started to promote the use of patterns amongst the abstract and concrete user interface layers of the CAMELEON reference framework. Patterns for other core models, like dialog, presentation and layout patterns have been introduced in different MBUID processes. In [19] the *User Interface Pattern Extensible Markup Language* (UsiPXML) has been introduced that uses the PLML pattern form combined with UsiXML models to specify and interlink patterns from task, dialog, presentation and layout model. Furthermore, a wizard plug-in for Eclipse has been developed, that guides the users throughout the pattern application process. The *Interface Development Environment for AppLications specified in UsiXML* (IdealXML) [20] uses patterns to exchange design experience amongst different developers and to increase reuse in a model-based development process. A pattern is represented by one of four core user interface models (task, mapping, presentation and domain model. Furthermore, an extended PLML form has been introduced which has been enriched with elements that support a SWOT analysis for evaluating patterns. In [21, 22] the *Pattern-based Modeling and Generation of Interactive Systems* (PaMGIS) Framework has been introduced, which addresses the model-based development of interactive systems. In PaMGIS several models (task, user, device and context models) are used to generate code for different target languages – e.g. C++, C#, Java and HTML. PaMGIS introduces an explicit "automation" element to enable the automatic code generation.

While these approaches showed how patterns can be integrated into MBUID processes, they suffer from different shortcomings. IdealXML for example doesn't support the user in the pattern integration process. The tool whether supports the specification of pattern parameters neither does it allow application of a pattern on a given problem context. Therefore patterns in IdealXML represent a rather low generativity and only serve as static model fragments, which have to be manually adapted by the developers. Other approaches, like PaMGIS, use a concept of constraints that supports the specification of pattern parameters, which can be used to derive problem specific pattern instances. But since PaMGIS implicitly includes layout information within the automation element (by the relation and anchor element) of the pattern language it doesn't support a clear separation of concerns. Furthermore, this intermixing results in a very fine-grained (and complex) pattern language, which conflicts with the idea that patterns represent a "bigger picture" [23].

UsiPXML represents one of the more mature approaches, since it encapsulates the core models from the pattern specification and allows the definition of pattern parameters for a flexible instantiation. Although UsiPXML interlinks patterns on different abstraction levels, these links are rather static. This hinders the flexible selection of pattern alternatives and limits the mightiness of the pattern concept. Based on these observations we state, that there is a demand to further investigate how patterns can be formalized and used to increase the flexible reuse of different model components within a user-centered MBUID process. In the following sections we present a first approach of a concept that addresses the before mentioned shortcomings of existing pattern-driven engineering approaches.

3 Towards a Framework for Pattern-Driven Engineering

In this section we present the three core user interface models of our MBUID approach that are most relevant for our pattern framework which is presented afterwards.

First, a *Room-based Use Model* (RUM) [24] is used to specify the user's tasks and goals in a platform independent task model. To separate the tasks ("what" is displayed) from the design aspects ("how" the tasks are presented) we have introduced an explicit presentation model and mapping language [5]. This presentation model is specified in the *User Interface Markup Language* (UIML) [25] and specifies the concrete interaction objects as well as their layout on the graphical user interface. Additionally, we started to incorporate the *Dialog and Interface Specification Language* (DISL) [26] for specifying the user interface behavior in a separated dialog model. While the RUM is used for structuring the users' tasks in the conceptual design phase, dialog models represent finite state machines that are interpreted during run-time, representing the executable tasks for the current situation. This further improves the separation of concerns and closes the gap between the RUM and the modality-dependent – but still platform-independent – presentation model.

For promoting a better scalability and a higher degree of reuse we identified patterns as a promising solution that can help to reduce the complexity of recurring design tasks. In Fig. 1 the initial version of our pattern framework for our development process is introduced.

A pattern language has been developed to organize and manage the user interface patterns and the model artifacts they represent. Overall the language capsulates and links patterns on three different levels of abstraction. Task patterns represent the highest form of abstraction. They focus on formalizing recurring user tasks – e.g. maintenance steps for a field device. The solution of these patterns is modeled within a RUM artifact, specified in the *Useware Markup Language* (useML) 3.0 [27]. Dialog patterns are specified in DISL and used to capture recurring navigational user interface aspects. Presentation pattern have the most concrete character in our framework presenting the layout aspects as well the use of concrete interaction objects. Presentation patterns are modality-dependent and formalized with UIML model fragments.

To fully employ the power of the language and to drive the development process, the introduced patterns are interlinked via semantic relationships within the pattern language. Due to these relationships the patterns can be further categorized. *Single Patterns* represent abstract building blocks in one modeling language that can be used by other patterns. For example the Split-Pane presentation pattern is such a single pattern in our language, which is reused in the Extras-On-Demand pattern (see Fig. 3).

Intra-Model Patterns are more constructive then Single patterns since they use lower patterns within the same modeling language for specifying the solution. In the same time these patterns are less abstract since they represent a refinement of other patterns. For example the Overview Detail pattern can be considered as an intra-model pattern, since it uses an instance of the Split-Pane pattern within its solution specification.

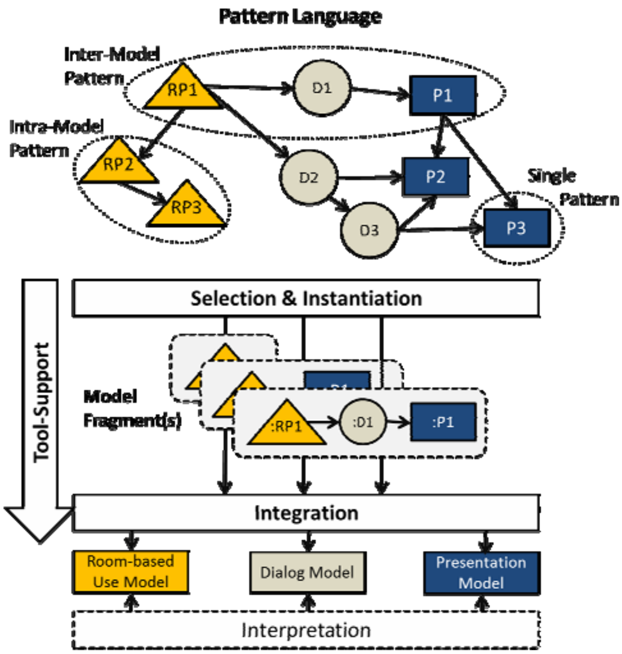


Fig. 1. The Pattern Application Framework

Inter-Model Patterns are the most powerful patterns in our framework. Besides the reuse of other patterns in the same modeling language their solution capsulates user interface aspects which are described in more than one modeling language. They also can be interlinked with patterns of other user interface modeling languages. Inter-Model Patterns may also further be interlinked with other Inter-Model Patterns that have a higher degree of abstraction.

To integrate the pattern language into the overall development process a sub process for the pattern application is demanded. Like proposed in [6, 28, 29] we specify four sequential core phases that have to be implemented by the pattern application tools: The pattern consumer starts with the (implicit) *Identification* of the model subset which should be refined by a pattern. Within the *Selection* step the pattern for the current problem context is selected. After that a concrete pattern instance is derived in the *Instantiation* phase by passing problem context specific pattern parameters. The last process phase is represented by the Integration of the pattern instance into the identified application spot of the target model.

In the next section we introduce our concept of a pattern form which enables the definition of a user interface pattern and supports the implementation of a pattern application process.

4 Towards a Pattern Notation for Generative Patterns

Generative pattern forms are demanded for a formal description of a patterns' solution. In Fig. 2 the structure of a generative pattern solution is proposed. Each pattern solution consists of three core sections. The *Model Fragments* describe the pattern solution in form of at least one model instance specified in one of our core modeling languages useML, DISL or UIML. Since patterns represent building blocks, these model fragments have to be flexible so that they can be instantiated in a given problem context. To support this without touching the language specification we use a concept of unique identifiers to address model elements as well as variables for implementing flexible attribute values in the model fragment. Addressing model elements is important to design patterns that support the specification of flexible substructures. For instance, in case of the Split Pane pattern, the pattern consumer should have the possibility to instantiate the pattern with a varying number of sub panes, which are modeled as containers in UIML.

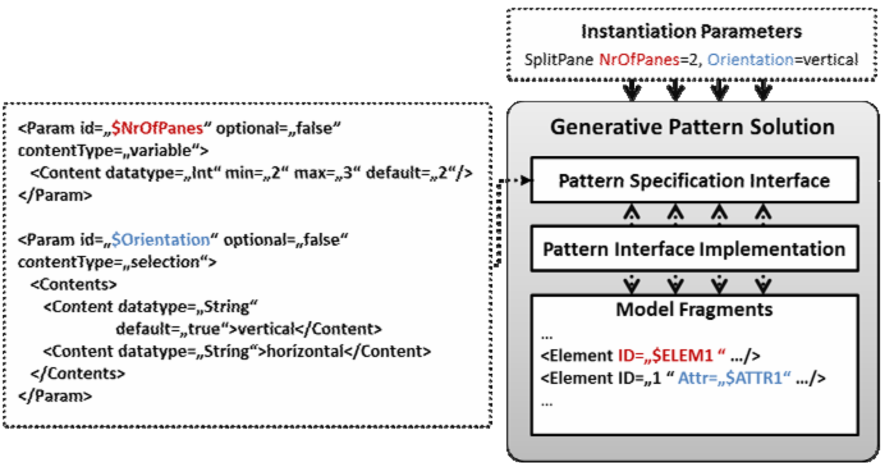


Fig. 2. Concept of a generative pattern form exemplified on the Split Pane pattern

Although it is important to give pattern consumers the possibility in adding/removing substructures to models and assigning values this also imposes a problem if the consumer misinterprets the pattern. Specifying constraints can be helpful to precisely describe which instances of a pattern are valid. While this might limit the degree of variability of a pattern it reduces the danger of misinterpreting and misusing a pattern The *Pattern Specification Interface* is used to exploit and constrain the patterns variables that can be set during the instantiation phase by the pattern consumer. As exemplified in Fig. 2 the pattern interface is specified via parameters. As introduced in [19] a parameter represents one external available variable that has an effect on the pattern model elements or attributes. A parameter has a unique name and an optional short description to document the parameters semantic. Following content types can be specified by a parameter:

- **Variable:** A variable represents a single data value used for the specification of standard parameters like texts for interaction objects. A variable is coded in one of the simple XML data types (e.g. String, Integer…). Furthermore, variables allow the specification of a lower and upper boundary (min, max) as well as a default value. In Fig. 2 the parameter "$NrOfPanes" is implemented as a variable, which specifies that a Split Pane pattern can have 2 to 3 panels for representing the content.
- **Selection:** A selection represents a set of constant data values and is used for the specification of fixed values like attribute lists. Like variables, selections use the content element to specify a list of parameter values. In Fig. 2 the orientation of the Split Pane pattern instance is specified as an parameter selection. The pattern consumer can chose between a vertical or horizontal orientation of the content containers, when instantiating the pattern.

Based on the pattern specification interface the *Pattern Interface Implementation* section is used to specify the according model manipulation actions during the instantiation phase of a pattern. Four basic model manipulations are used to tailor the generic model fragment according to the instantiation parameter values. The *add*-, *remove*- and *replace*-operations are needed for manipulating sub structures within the pattern model fragments. These operations are used for example to insert sub containers within the Split-Pane pattern or to add concrete interaction objects in case of the Main-Menu pattern. An *assign*-operation is used to pass parameters of the pattern interface to the model fragments and to assign the parameter values during the pattern instantiation. For example, the Orientation attribute value of the Split Pane pattern is assigned during the pattern instantiation to the according UIML attribute.

We use XSLT as the implementation language for the Pattern Interface Implementation since the language specification supports the before mentioned required operations.

5 Illustration of the Pattern Framework

In the presented scenario, patterns are used to develop a user login screen. The design process starts with the specification of the users task in useML. The Login Form pattern is a intra-model pattern that specifies the basic tasks of *Enter Username*, *Enter Password* and *Login* in the RUM and refines them in the presentation model by mapping these tasks on concrete interaction objects. Parameters are used in this pattern to assign the descriptions of the interaction object labels. This inter-model pattern further implements different other patterns in its presentation model. For example the Extras-On-Demand pattern is used for the specification of the *Remember Me* and *I forgot my password* interactions, which are presented on demand. The Extras-On-Demand pattern itself implements a vertical Split Pane pattern with two panes separating the Login-Form from the Extras-on-Demand content.

After the task model specification the user interface layout is specified. In the first step the user interface is split into two main panels, one for navigation and one for presenting the content. To do so, an instance of the Split Pane pattern is derived with two panes and horizontal alignemt specified via the patterns' parameters. Each content pane of the split pane pattern has a source which can be further refined by other presentation patterns or user specified UIML content. After that the left panes' source is refined by applying the Main Menu pattern. This pattern is a inter-model pattern that adds navigational aspects to the dialog model. For each of the three navigation links that have been passed via the patterns' parameters an abstract interaction object is added to the dialog model. These objects represent the triggers for the navigation transitions in the DISL model. After that, the UIML fragment is instantiated and the according concrete interaction objects as well as the events for triggering the DISL transitions are generated.

The output of the pattern application process is a set of models that describes the users' login task, the user interface transitions to other dialog states and the presentation of the login screen. The resulting rendered user interface is depicted on the right side of Fig. 3. After the instantiation these models have to be further refined by the developer. For example it has to be specified, what action is executed when the user clicks the login button.

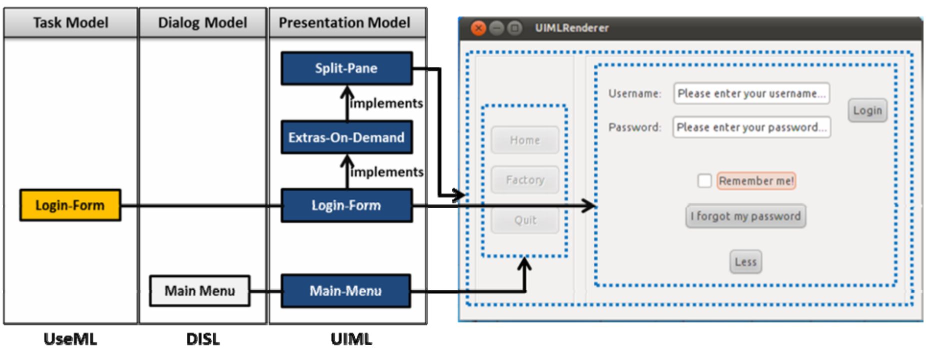


Fig. 3. Inter-Model Patterns (left) and the Resulting Rendered UI (right)

6 Summary and Outlook

In this paper we presented a concept for the pattern integration into our MBUID process. A framework has been introduced that defines different types of patterns – single, intra-model and inter-model patterns – that represent different levels of abstraction in the pattern language. To promote the separation of concerns we use different models for the implementation of a pattern. A concept of a generative pattern form has been introduced that enables the specification of pattern instance components. The pattern specification interface and pattern interface implementation supports the fine grained definition of valid pattern model instances without touching the core models' language specifications.

Since our work presents a research in progress we have to address several aspects in the future. Tools will be developed that support the pattern consumers from the identification to the integration phase of the pattern application process. Another crucial aspect is to refine the inter-model pattern relationships. A modular linking of patterns on different abstraction levels promises to improve the quality of the pattern language since the pattern consumers can search for refined patterns that aggregate other pattern instances and may better fit on the given problem context. Besides these tasks we want to continuously extend the pattern language with new patterns to test if the pattern notation itself is sufficient to formalize the identified patterns and further to extend the set of reusable user interface artifacts.

Acknowledgments. Our special thanks go to Peter Reuter and Markus Kleine who helped us with the design of the initial pattern framework and the implementation of first prototypical tools for testing the pattern application process.

References

1. Breiner, K., Görlich, D., Maschino, O., Meixner, G., Zühlke, D.: Run-Time Adaptation of a Universal User Interface for Ambient Intelligent Production Environments. In: Proceedings of the 13th International Conference on Human-Computer Interaction. Part IV: Interacting in Various Application Domains, pp. 663–672. Springer, San Diego (2009)
2. Seissler, M., Meixner, G., Breiner, K.: Using HCI Patterns within the Model-Based Development of Run-Time Adaptive User Interfaces. In: Proceedings of the 11th IFAC/IFIPS/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France (2010)
3. Zuehlke, D.: SmartFactory - Towards a factory-of-things. Annual Reviews in Control 34, 129–138 (2010)
4. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for Multi-Target User Interfaces 15, 289–308 (2003)
5. Gaffar, A., Seffah, A., Poll, J.A.V.D.: HCI pattern semantics in XML: a pragmatic approach. In: Proceedings of the 2005 Workshop on Human and Social Factors of Software Engineering, pp. 1–7. ACM, St. Louis (2005)
6. Sinnig, D., Forbrig, P., Seffah, A.: Patterns in Model-Based Development. Position Paper in INTERACT 2003 Workshop entitled: Software and Usability Cross-Pollination: The Role of Usability Patterns (2003)
7. Kotzé, P., Renaud, K., Koukouletsos, K., Khazaei, A.D.: Patterns, Anti-Patterns and Guidelines – Effective Aids to Teaching HCI Principles? In: Proceedings of HCIEd 2006-1, 115–120. University of Limerick, Limerick (2006)
8. Winn, T., Calder, P.: Is this a pattern? IEEE Software 19, 59–66 (2002)
9. Alexander, C.: The Timeless Way of Building. Oxford University Press, Oxford (1979)
10. Hennipman, E., Oppelaar, E., van der Veer, G.: Pattern Languages as Tool for Discount Usability Engineering. In: Interactive Systems. Design, Specification, and Verification, pp. 108–120 (2008)
11. Alexander, C., Ishikawa, S., Silverstein, M.: A Pattern Language: Towns, Buildings, Construction. Oxford University Press, Oxford (1977)
12. Gamma, E., Helm, R., Johnson, R.E.: Design Patterns. Elements of Reusable Object-Oriented Software. Addison-Wesley Longman, Amsterdam (1995)
13. Tidwell, J.: Common Ground, `http://www.mit.edu/~jtidwell/interaction_patterns.html`
14. Borchers, J.O.: CHI meets PLoP: an interaction patterns workshop. SIGCHI Bull. 32, 9–12 (2000)
15. Vanderdonckt, J., Simarro, F.M.: Generative pattern-based design of user interfaces. In: Proceedings of the 1st International Workshop on Pattern-Driven Engineering of Interactive Computing Systems - PEICS 2010, Berlin, Germany, pp. 12–19 (2010)

16. Wolff, A., Forbrig, P.: Pattern Catalogs using the Pattern Language Meta Language. In: Proc. of the Workshop Visual Formalisms for Patterns at VL/HCC 2009 (2009)
17. Paternò, F.: Model-based design and evaluation of interactive applications. Springer, London (1999)
18. Gaffar, A., Sinnig, D., Seffah, A., Forbrig, P.: Modeling patterns for task models. In: Proceedings of the 3rd Annual Conference on Task Models and Diagrams - TAMODIA 2004, Prague, Czech Republic (2004)
19. Radeke, F.: Patterns in Model-based Design of User-Interfaces: Towards an Integrated User-Interface Development Approach. Vdm Verlag Dr. Müller (2008)
20. Montero, F., Lozano, M., González, P., De Investigación Louise, G.: IdealXML: an Experience-Based Environment for User Interface Design and pattern manipulation (2005)
21. Engel, J., Märtin, C.: PaMGIS: A Framework for Pattern-Based Modeling and Generation of Interactive Systems. In: Jacko, J.A. (ed.) HCII 2009, Part II. LNCS, vol. 5610, pp. 826–835. Springer, Heidelberg (2009)
22. Engel, J., Märtin, C., Forbrig, P.: Tool-support for pattern-based generation of user interfaces. In: Proceedings of the 1st International Workshop on Pattern-Driven Engineering of Interactive Computing Systems - PEICS 2010, Berlin, Germany, pp. 24–27 (2010)
23. Dearden, A., Finlay, J.: Pattern Languages in HCI: A Critical Review. In: Human–Computer Interaction. pp. 21–49 (2006)
24. Görlich, D., Breiner, K.: Intelligent Task-oriented User Interfaces in Production Environments. In: Workshop: Model-Driven User-Centric Design & Engineering, Seoul (2007)
25. Abrams, M., Phanouriou, C., Batongbacal, A.L., Williams, S.M., Shuster, J.E.: UIML: An Appliance-Independent XML User Interface Language, vol. 31, pp. 1695–1708 (1999)
26. Schäfer, R.: Model-Based Development of Multimodal and Multi-Device User Interfaces in Context-Aware Environments. Shaker Verlag, Aachen (2007)
27. Görlich, D., Breiner, K.: Useware Modeling for Ambient Intelligent Production Environments. MDDAUI (2007)
28. Radeke, F., Forbrig, P.: Patterns in Task-Based Modeling of User Interfaces. In: Winckler, M., Johnson, H. (eds.) TAMODIA 2007. LNCS, vol. 4849, pp. 184–197. Springer, Heidelberg (2007)
29. Sinnig, D., Javahery, H., Forbrig, P., Seffah, A.: The Complicity of Model-Based Approaches and Patterns for UI Engineering. In: Proceedings of BIR, pp. 120–131. Shaker Verlag, Ithaca (2003)

Complex Components Abstraction in Graphical User Interfaces

Pedro M. Teixeira-Faria and Javier Rodeiro Iglesias

School of Technology and Management - Polytechnic Institute of Viana do Castelo, Portugal
pfaria@estg.ipvc.pt
School of Informatics Engineering, University of Vigo, Spain
jrodeiro@uvigo.es

Abstract. This paper proposes a structure to increase the level of abstraction in building visual user interfaces, establishing an initial set of features and properties that a system for abstract complex components specification must have. Two approaches of AIOs have been studied: Interactors and ADV. The study was focused on the way in which these techniques manage interaction and presentation of interface objects. In the following section an interface complex component definition is proposed. In section 3.1 a basic set of properties and features for interface complex components are showed. In the last section an analogy between classes and objects in object-oriented programming and interface complex components are presented.

Keywords: Complex Components, Abstract Interaction Objects.

1 Introduction

Some work in the field of interactive graphics involves describing the interface of a system in terms of a collection of interaction objects (components). This approach applies both to the level of implementation [12] (where the term *widget* is sometimes used), and at the more abstract level of system specification [11]. A user interface can be specified considering a framework formed by several *interface design steps* [13], represented with a proper user interface specification. Those steps correspond to models that describe different aspects of the user interface. For each model we can find diverse constructors in the literature [15] and different purposes. Following we indicate some of these constructors and their abbreviations for future reference. An *abstract interaction object* (*AIO*) represents a user interface object without any graphical representation and independent of any *environment* [15]. Usually, is understood as a conceptual representation of an interface object. A *view* is a collection of *AIO's* logically grouped to deal with the inputs and outputs of a task [15]. An *environment* is the *global interface* to which the interaction object belongs [6]. A *concrete interaction object* (*CIO*) represents any visible and manipulable user interface visual component that can be used to input/output information related to user's interactive task. *CIOs* are sometimes called *widgets* (*wi*ndows ga*dgets*) or *physical interactors*, which often include simple objects (e.g. edit box, push button)

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 309–318, 2011.
© Springer-Verlag Berlin Heidelberg 2011

[4]. [13] uses the *CIO* terminology with other significance: *conceptual interaction object*, which has multiple states and with which is possible to interact. This is the most close with the *complex component* concept which will be introduced on this paper.

Most of the visual interfaces are created for the user to interact with them. However, the visual components of an interface can be divided into *interactive* (with one or multiples states and behaviors) or *no interactive* (visible, but without related events, basically having just one state). The main objective of this paper focuses on interactive visual components, meaning those which presuppose the existence of *dialog*. Much work in the field of interactive graphics involves describing an interface in terms of a collection of *interaction objects* [6][8][14]. [9] refers that the notion of *interaction object* do not need to be confined to graphics systems, because it represents a useful structure for thinking and reasoning about the behavior of interactive systems in general. Usually, *interaction objects* are seen as an independent entity with a local state that can engage in events with its *global interface*, possibly resulting in changes to the state. Actually, *interaction objects* are not necessarily visual, but *no visual interaction objects* are a weak percentage of the interaction objects total. The following section presents two concepts that increase the level of abstraction on the *AIOs* specification. The next section focuses on presenting the concept of *complex component*, establishing the features that a *complex component* must have, referencing also union mechanisms between *complex components* and the *global interface* in which it is represented. Then some similarities between *complex components* and *object-oriented paradigm* are presented. On last section some conclusions are taken and future work is envisaged.

2 Abstract Interaction Objects

Each human computer interaction relies on three dimensions: the *user*, the *task* and the *environment*. Thus, on the interface designing process three steps can be identified: (1) the selection of appropriate *AIOs*; (2) the transformation of *AIOs* into *CIOs* and (3) the placement of *CIOs* to form the final observable *environment* [3]. On this section we introduce two concepts related with the above step (1): *interactors* and *abstract data views*.

2.1 Interactors

[6] introduced the concept of *interactors* which are components (objects) in the description of an interactive system. With *interactors* is possible to structure interactive systems models using the concept of an object that is able to present part of his state to outside of him. Two models have adopted this concept on their specifications: CNUCE and York [9] [10]. The formalisms used to express these models afford different approaches to the construction and analysis of specifications. On the case of CNUCE model, at a very abstract level, an *interactor* is viewed (externally) as what the authors called a "*black box*" that mediates a "*user side*" and an "*application side*". It can receive information from either side, process that information and return it.

The *interactor* model developed at York [6] is based on *states*, *commands*, *events* and *renderings*. Thus, at an abstract level, an *interactor* encapsulates a *state* (defined as a set of attributes) which is reflected through a *rendering relation* (*rho*) onto some perceivable *representation* (*P*) [9]. Typically, an *interactor* has a specific functionality and a unique appearance of his state, because models do not consider multiple *rendering functions* for the state of an *interactor*. This is because the model specification is based on the *interactor dialog* rather than in its *visual appearance*. The interface between an *interactor* and its *environment* consists of an event set (defined as a set of actions) of two kinds: *responses* (events generated by the *interactor*) and *stimuli* (events that reach the *interactor* from the *environment*). Is important to note that an *interactor* is an algebraic conceptualization and as such its meaning does not contemplate the existence of a visual presentation. Never is called a function of a visual renderer, but only an algebraic renderer. And so, an *interactor* is viewed more as a useful concept that can be used in formal analysis of an interactive system.

2.2 Abstract Data Views (ADVs)

[1] refers the increasing number of authors advocating the concept of dialog independence, where interactive systems are designed and implemented with the goal of providing a clear separation between the user interface and the application. Many advantages of dialog independence referred by the authors are, for example: 1) ease of maintenance of the existing system; 2) support for alternative visualizations of a single system; 3) design and application reuse.

A user interface design concept which corresponds to the characteristics mentioned is the *Abstract Data View (ADV)*. *ADVs* are used as an interface between the computer system and the user and have been viewed as a design paradigm for user interfaces [7]. An ADV conforms to the state of its associated *abstract data object* (*ADO*) [2]. Different *ADVs* may view the same *ADO*, since interfaces can support alternate "*views*" of data (or modes of interaction) [2]. In order to maintain separation of concerns and promote reuse, *ADOs* are specified to have no knowledge of their associated *ADVs*. The connection between the *ADO* instance and *ADV* instance captures the *WYSIWYG* nature of a user interface [7]. At first glance this design would appear to be the MVC model [5]. However, in MVC the controller handles input, whereas in the *ADV* model the input and the output are handled by each *ADV* instance, and vertical consistency is maintained by the objects managing the relation between the *ADV* and *ADO* instances. Both *ADVs* and *ADOs* can be used upon by actions to change or query their state. Those actions can be divided into two categories: *causal actions* and *effectual actions*. *Causal actions* denote the input events acting on an *ADV* when it is acting as a user interface or interface to some other media. These actions are triggered from outside the system and internal objects are not able to generate this type of action. A "*keystroke*" or a "*mouse click*" are simple examples of input events that are *causal actions*. *Effectual actions* are the actions generated directly (or indirectly by a *causal action*) and are supported by both the *ADVs* and *ADOs*. The triggering of an *effectual action* by another action will normally be a synchronous process. This type of actions can be viewed as the activation or call of a method or procedure that is part of the public interface of an *ADO* or *ADV*. It is possible to indentify similar features between *interactors* and *ADVs*, concerning the possibility of both receive/send events from/to the *environment*.

Table 1. *Interactors* and *ADVs* input/output

	Receives (in)	***Sends (out)***
interactor	*stimuli*	*response*
ADV	*causal action*	*effectual action*

Is important to refer that *interactors* and *ADVs* makes a description of interface objects but do not allow describing the complete interface. They only allow defining the interaction between the interface objects.

3 Complex Components

Until now, were presented two interface specifications that consider *AIO*: *interactors* and *ADVs components*. However, the objective is intended to seek a specification that allows the definition of visual interface components, their composition and interaction (*dialog*) on these same components. In the case of *interactors*, they appear as an algebraic conceptualization whose usefulness will be higher in the formal analysis of an interactive system. *ADVs* are closely linked with the concept of *ADO*, and act as the views of those objects. Both components, *interactors* and *ADVs*, do not consider their specification in the context of a complete interface, focusing primarily on defining simple components with multiple states and the relationship between them. A new concept is presented here and is called *complex component*, which should be integrated within a specification that fits the parameters of interface components definition, components composition and user interaction over those components.

We have introduced the concept of interface component as appears in literature, showing different interface components and their properties. As it was possible to verify, each component is individually defined and can make relationship with other components on the interface. Now, we will introduce the concept of *complex component* and then we detail their properties. Basically, a *complex component* is a component composed of other components (*simple* or/and *complex*) interact with each other through its *self events* (described below) working toward a common goal (e.g. a toolbar allows a user to select a specific tool to perform a task at a given time). The components follow a hierarchical topological structure, so that each can be contained within others. The idea of representing a *complex component* has the underlying need of existing a component container (not necessarily visual) which, together with other visual *simple* (or/and *complex*) components will constitute a *complex component.* Besides acting as a container, a *complex component* also has its own states and the ability to interact with other *simple* and *complex components*, external to him.

A *complex component* allows the possibility to define different spaces of activation (interaction) especially on the visual components in direct manipulation interfaces. With regard to its importance for these components, the areas must have visual significance: those areas must be visual components. Within *complex components* may exist those with one activation area and those which have more than one activation area. An example of those which have a single activation area is the button (can have several states). Usually, the user interacts in the same physical area. An example of a component with more than one activation area is a toolbar that usually has several areas of interaction (options) with different spatial positions. Thus, this is an important issue to be studied on a *complex component*: which have more than one state and more than one activation area. One important issue related with the *complex component* abstract definition is his close relation with the *abstract composition model* [13] since it is necessary to establish spatial rules between the components inside the container, and the spatial rules of him related with the *global interface* on which it appears. Is important to state that *complex components* always have dialog, otherwise we are only in the presence of components *topological composition* and *visual appearance*. There are several advantages in creating the *complex component* concept, highlighting four of those advantages:

- Reduces complexity of user interface design;
- Allows reuse of components;
- Allows formal specification of UI parts;
- Is near to the user tasks model by tasks decomposition.

The use of *complex components* to specify a user interface will allow entropy decreasing, resulting that the complexity of the interface decreased, by the fact that this component type increases the specification abstraction level. This happens because the interface creation problem can be subdivided in small parts represented by complex components. Another advantage that *complex components* have is the ability to reuse components in the design of the user interface. A *complex component* is not dependent on the interface where is created, which means that due to its flexibility it is possible to reuse it in different interfaces. This advantage of reusing components simplifies the interface specification. For example, this allows reusing components with certain behavior, to make visual changes and still maintain its functionality. Another advantage is that using *complex components* is possible to formally specify each component individually and then, at the *global interface* level, is only necessary to formally specify the part that is not present in the *complex components*. Thus, it is possible to define *complex components* separately, is also possible to verify them separately. The last advantage indicated above concerns the *complex component* architecture, which has an approach to the task model, in sense that it performs task decomposition. This analogy happens since each component can be decomposed into a hierarchy of components similar to a hierarchy of objects with their specific properties.

3.1 Complex Components Features

We establish several features that a *complex component* must have. One is progressive level of abstraction (from basic to more complex specification) at three different aspects:

- Visual appearance;
- Topological composition;
- Interaction.

The above three aspects can be related to some of the *interface design steps* [13]. The *visual appearance* representation usually is presented on the *abstract presentation model*, which includes, for each component, his visual description. The *topological composition* appears represented on the *abstract composition model*, responsible for defining the spatial relations between components. Usually, the interface components *dialog* is established in the *dialog model*.

It is possible to verify on (Fig. 1) the typical structure of a *complex component*, structured in the three abstraction levels considered on his definition. Observing that figure we can infer that a *complex component* manages the *visual appearance* and *topological composition* of the components (*simple* or/and *complex*) that are inside of it, due to its characteristics of component container. Furthermore, from that figure we can see that the *complex components* consider *interaction*.

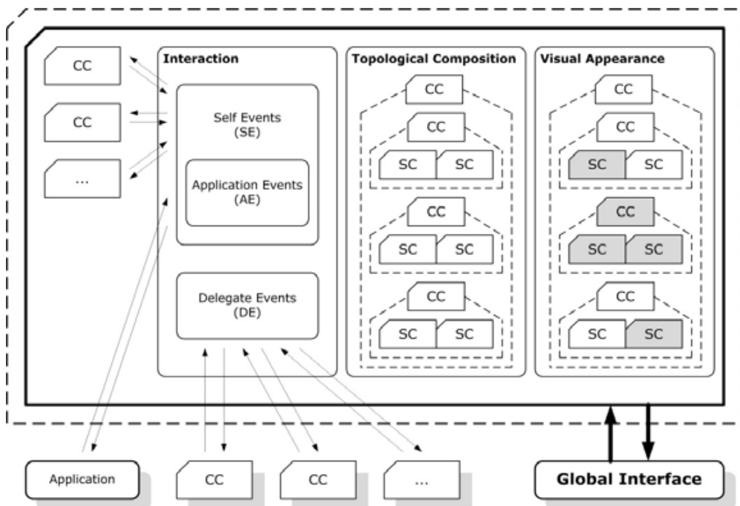


Fig. 1. A *complex component* typical structure

Visual Appearance. (Fig. 1) represents a *complex component* (bounded by a ticker black rectangle) which corresponds to the (*CC*) on the top of the *visual appearance* rectangle with rounded corners. This *complex component* is composed of three other *complex components* which in turn are composed by two *simple components* (*SC*). A *simple/complex component* has several properties [13], one being *visibility* (*visible* or *invisible*). For example, in figure, the *simple/complex components* represented in gray refer to components with the property *invisible* active. The other *simple/complex components* represented in white have the *visible* property active. That means, on a real *global interface*, which components are visible at a given time.

Topological Composition. From (Fig. 1) is easy to verify that *complex components* consider topological composition. The *complex components* follow a hierarchical topological structure, so that each can be contained within others. In the figure is illustrated the *topological composition* of four *complex components* and six *simple components*, arranged on a three levels hierarchy. Inside *topological composition* rectangle with rounded corners are represented several *simple/complex components*. The (*CC*) on the top corresponds to the first hierarchy level.

Interaction. *Interaction* is included in *complex components* and occurs between them and between them and the *global interface* to which they belong. In addition to *topological composition* and *visual appearance*, a *complex component* will manage (to trigger or to receive) two kinds of events: *self events* and *delegate events*.

- *Self events (SE)*: on a *complex component* there are visual transitions between components (*simple* or/and *complex*) inside of that *complex component*, that affect only the *complex component* and not the components that are outside of him. These transitions are triggered by (*SE*) which are events that proceed from a component that is within the same *complex component* which receives it. With respect to establishing dialog with the application, a *complex component* also has mechanisms of interaction in both ways through (*SE*) (*application events* (*AE*) are represented in (**Fig. 1**) as a subpart of the *self events*) which means that the *complex component* can sends/receives events to/from the application.
- *Delegate events (DE)*: is the other event type which proceeds from a component that is outside of *complex component* which receives it (meaning that events over a *complex component* may cause changes in other interface components outside of it (represented by the dashed rectangle)).

Global Interface Component. *Interaction* can happen between *complex components* and between *complex components* and the *global interface*. The global interface can be represented by a *complex component* and is responsible for the union between components at the higher topological level on the interface.

3.2 Similarities between Complex Components and OO Paradigm

The *object-oriented programming* (*OOP*) paradigm is quite widespread and its robustness has been validated by the vast community of programmers who use it. The *complex component* concept here described is an abstraction which represents a simplified description of a system (a user interface visual component on this case) which captures the essential elements of that system. In (*OOP*) the abstractions are mostly represented as classes. Classes support *encapsulation*, *inheritance* and *polymorphism*. They are mostly static and define the capabilities of an object (which is a dynamic entity) and every object has *identity*, *state* and *behavior*. It is possible to find similarities to (*OOP*) paradigm, establishing an association between *complex component* features and *class* concepts (Table 2).

Table 2. Similarities between *Complex Components* and *OO Paradigm*

OO Paradigm	*Complex Component*
encapsulation	*modularity and information hiding*
inheritance	*features and events reuse and addition of new ones*
polymorphism	*allows visual presentation changes keeping functionality*
identity	*each one has its own and unique identifier*
state	*a visual appearance at a given moment that waits for an event from the user or from the application*
behavior	*the behavior is represented by events on the complex component dialog*

Encapsulation. *Encapsulation* is a mechanism that brings together the code (*methods*) and the data (*attributes*) keeping them controlled in relation on their access level. One important feature envisaged by *complex components* is his *encapsulation* capability emerging two advantages from that:

- *modularity*: *complex components* can be reused on other interfaces, allowing changing its *visual appearance* and to keep its *behavior*. This feature keeps the system flexible and becomes easier to have a better description and organization of used components, which allows for greater ease in building and maintaining the global user interface.
- *information hiding*: the possibility to define public and private information is important to establish communication rules between *complex components*, concerning his capability to trigger and respond to *self events* (*SE*) and *delegate events* (*DE*) (which may be related to *public* and *private* methods of a class).

More important becomes this *encapsulation* feature in a *complex component*, since it contains composition at three levels (*visual*, *topological* and *dialog*). From the point of view of *complex component* features, is more difficult to do *dialog composition* characterization than the other two compositions. This happens by the fact of the interactive nature of a *complex component*.

Inheritance. *Inheritance* is the ability to specialize classes (*object types*) so that new classes contain (in addition to structural and behavioral characteristics as defined by their "ancestors") other characteristics set for them own. When *complex component* is created (topologically above other(s) that already exists) features and behavior (events) are inherited, being possible to add new features and behaviors. This means functionality reuse, which increases interface *complex components* organization when they are being created.

Polymorphism. *Polymorphism* is a term derived from Greek and means many forms. In (*OOP*) means the possibility of an abstract type (interface or abstract class) to be used to create multiple objects, reusing the methods. [15] uses *polymorphism* on virtual interaction objects creation, allowing plural objects instantiations from different *application programming interfaces* (*APIs*) manipulated via the same abstract object instance. On a *complex component* the *polymorphism* occurs when, for example, we want to modify the graphics that make up the *visual appearance*. This can be done without the need to change *complex component* functionality.

After presentation of the previous three properties that *complex components* have in their composition, objects created when the interface is being executed have three other properties that are presented below.

Identity. The *identity* is the property of a thing which distinguishes it from all other objects (is unique and identifiable). A *complex component* has it, which guarantees that his identifier is unique and is not used on any other component.

State. *State* is the set of values that an object encapsulates (two objects may have identical *state*, and yet are still separate, distinct, identifiable objects). On a *complex component*, distinct visual appearances represent different *states*.

Behavior. *Behavior* is the set of actions that an object knows how to take (usually known as *methods*). The dialog on *complex components* results from *self events* (*SE*) and *delegate events* (*DE*) derived from user or/and application *interaction*.

4 Conclusions

When referring *interaction* we focus its importance in the role of user in it. However, with this work we are focused on the interaction between abstract interface *complex components* when the user interacts with the interface. We can consider an interface as a set of interface simple components that user perceives or as a set of interface *complex components* that have the same goal and the user perceives as a unique entity of user interface. We showed that the *complex component* concept makes sense if considered integrated with other components to create the global user interface. As more progressive and higher is the abstraction level of a *complex component* simpler becomes a visual user interface representation. The definition of *complex components* seems to indicate that it is possible to use them to facilitate user interfaces creation, due to high level of abstraction that they introduce to build a complete representation of user interface. The study presented on this paper was made with the objective of identify the main features that an abstraction of *complex components* must include to build visual user interfaces, establishing the foundations for a future specification of *complex component*.

Acknowledgments. This work was supported by:

1) Grant SFRH/PROTEC/49496/2009 of MCTES – Ministério da Ciência, Tecnologia e Ensino Superior (Portugal)
2) Project TIN2009-14103-C03-03 of Ministerio de Ciencia e Innovación (Spain)
3) Project 10DPI305002PR of Xunta de Galicia (Spain)

References

1. Alencar, P., Cowan, D., Germán, D., Lichtner, K., Lucena, C., Nova, L.: A Formal Approach to Design Pattern Definition & Application. Technical Report CS-95-34, University of Waterloo, Waterloo, Ontario, Canada (1995)
2. Alencar, P., Cowan, D., Lucena, C.: A Logical Theory of Interfaces and Objects. IEEE Transactions on Software Engineering 28(6), 548–575 (2002)
3. Bodart, F., Noirhomme, M., Vanderdonckt, J.: Guidelines for Choosing Interaction Objects. In: VCHCI 1993, pp. 431–432 (1993)
4. Bodart, F., Vanderdonckt, J.: Widget Standardisation through Abstract Interaction Objects. Institut d'Informatique, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium (1996)
5. Carneiro, L., Coffin, M., Cowan, D., Lucena, C.: User Interface High-Order Architectural Models, University of Waterloo - Department of Computer Science (1993)
6. Carr, D.: Specification of Interface Interaction Objects. In: CHI 1994 – ACM Conference on Human Factors in Computer Systems, pp. 372–378 (1994)
7. Cowan, D., Lucena, C.: Abstract Data Views: An Interface Specification Concept to Enhance Design for Reuse. IEEE Transactions on Software Engineering (1995)

8. Duke, D., Harrison, M.: Abstract interaction objects. Computer Graphics Forum 12(3), 25–36 (1993)
9. Duke, D., Faconti, G., Harrison, M., Paternó, F.: Unifying views of interactors. In: Proceedings of the Workshop on Advanced Visual Interfaces, June 1-4, pp. 143–152 (1994)
10. Harrison, M., Duke, D.: A Review of Formalisms for Describing Interactive Behaviour. In: Taylor, R.N. (ed.) ICSE-WS 1994 and SE-HCI 1994. LNCS, vol. 896, pp. 49–75. Springer, Heidelberg (1995)
11. Jacob, R.: A Specification Language for Direct Manipulation User Interfaces. ACM Transactions on Graphics (TOG) 5, 283–317 (1986)
12. Myers, B.: A New Model for Handling Input. ACM Transactions on Information Systems 8, 289–320 (1990)
13. Rodeiro, J.: Representácion y Análisis de la Componente Visual de la Interfaz de Usuario, PhD Thesis, Universidad de Vigo (2001)
14. Savidis, A.: Supporting Virtual Interaction Objects with Polymorphic Platform Bindings in a User Interface Programming Language. In: Guelfi, N. (ed.) RISE 2004. LNCS, vol. 3475, pp. 11–22. Springer, Heidelberg (2005)
15. Silva, P.: User interface declarative models and development environments: A survey. In: Paternó, F. (ed.) DSV-IS 2000. LNCS, vol. 1946, pp. 207–226. Springer, Heidelberg (2001)

User-Oriented Accessibility Patterns for Smart Environments

Michael Zaki and Peter Forbrig

University of Rostock, Department of Computer Science,
Albert Einstein Str. 21, 18059 Rostock, Germany
{michael.zaky,peter.forbrig}@uni-rostock.de

Abstract. One of the main indicators concerning the usability of an application is the corresponding level of accessibility provided by this application. Although a lot of work has been done in the software engineering domain, the accessibility problem has not been enough tackled in the HCI area. In this paper we present an idea to resolve the user-related accessibility problems since the modeling stage of smart environment applications while being assisted by patterns. The proposed idea is to provide two generic patterns used for any accessibility modeling problem, and additionally the creation of two pattern libraries presenting concrete solutions for the most common user-based accessibility problems.

Keywords: Smart Environments**,** Bidirectional Information Accessibility, Task Models, Patterns.

1 Introduction and Motivation

The need for applications taking user accessibility problems into account has obviously increased over the last few decades. 1 out of 5 people in the United States has some kind of disability. The number of people with disabilities is still increasing, as it has increased with about 25 % in the last ten years. Consequently, the fact that an application is universally accessible has become one of the top concerns of this application's usability evaluation. A very common understanding of the word 'Accessibility' defines it as "*Authorization, opportunity or right to access records or retrieve information from an archive, computer system or a website*". In our point of view, the 'Accessibility' must be defined in a broader way while speaking about ubiquitous computing environments and especially smart environments for two main reasons:

a) In [1] smart environments have been defined as "*physical environments where actors are performing tasks cooperatively while being assisted by mobile and stationary devices*". From this definition, one can easily see that a given actor or user in a smart environment is receiving information not only from the existing devices and domain objects but also from the other actors, which implicitly means that in order to solve the accessibility problem we have to guarantee that this user is able to receive and understand the ambient information sent by the other actors as well as the devices and the computing systems in the environment.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 319–327, 2011.
© Springer-Verlag Berlin Heidelberg 2011

b) The user needs to perform tasks so that he can successfully accomplish his role and achieve the goal for which he is in the smart environment. But actually in order to accomplish his role, he may also need to deliver information to the other entities in the environments as actors, devices and computing systems. Consequently here we can see another dimension of the accessibility problem, as the user must also be able to deliver information to the environment. This new requirement increases the complexity of the accessibility problem, as we have to guarantee as well that the information provided by this user is accessible to any other entity in the environment.

Taking these requirements into account, we define the word ‘Accessibility’ in the domain of smart environments in the following manner:

“*The opportunity for all the users to receive and to deliver all kinds of information, regardless of the information format or the type of user impairment*”.And so here we are dealing with a need for bidirectional accessibility concerning any information coming from or going towards any user in the smart environment.In Fig.1 the information flow for the role listener in a presentation scenario is depicted, and the direction of the arrows represents the direction of the information. In this paper we present the solution of using patterns for the integration of the accessibility needs into the modeling process of smart environment applications.

The remainder of the paper is structured as follows: In Section 2, the related work and a background about the topic are presented. After that, the approach of using patterns to deal with the accessibility problems is discussed while in Section 4 the presented approach is fostered by an example which illustrates and puts into evidence the use of these patterns. Finally we summarize and give a brief overview of the future research avenues in this domain.

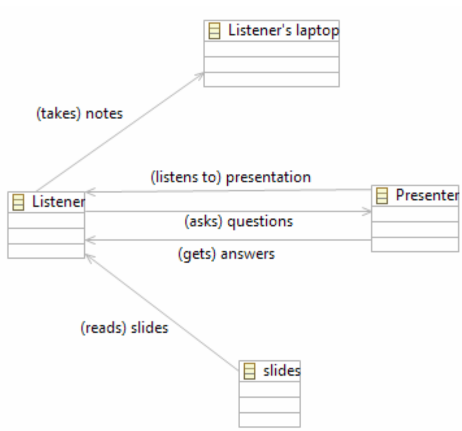


Fig. 1. Information flow for the role ‘Listener’ in a presentation scenario

2 Background and Related Work

In spite of the fact that patterns were first invented in urban architecture by Christopher Alexander in 1977[2], their influence has spread and reached the software engineering as well as the HCI area. In [3] The Gang of Four have introduced patterns as recurring solutions for common software design problems. These patterns realized a brilliant success in the software engineering field, and consequently the idea of using patterns has also conquered the HCI domain. In [4] Borchers defines HCI design patterns as "*a structured textual and graphical description of a proven solution to a recurring design problem*". Also in [5] Tidwell defines a collection of interaction design patterns where the solution is expressed in terms of suggested perceivable interaction behavior.

The accessibility is one of the fields where HCI patterns have already contributed. However, a lot of work needs to be done in this direction as the existing patterns are unable to deal with this problem in environments with high complexity like smart environments where the user is exchanging information with all the surrounding entities. For example in [6] accessibility patterns solving the web design and content accessibility problems were discussed, while in [7] an approach merging accessibility and usability UI needs by the use of patterns is presented. The main shortcomings are that these approaches are just coping with Human-web accessibility problems and they present solutions in order to make the application's UI accessible to every user, while in smart environments the user can receive information from other sources and consequently one can have other types of accessibility problems such as Human-human accessibility problems for example. An illustration of such an accessibility problem can be clarified if we imagine the case of a normal lecture taking place in a smart meeting room. If one of the existing students is having troubles to get information from the professor, so in fact this student is having a Human-human accessibility problem. Another major disadvantage is that the solution is usually presented on the UI design level, while it would be better and more effective if we can integrate a solution earlier in the application's modeling stage.

Task models proved a great success as a starting point for modeling applications in the HCI area. In [8] the idea of using task models for smart environment is tackled. The main argumentation why task models are suitable for these environments is the high complexity of task performance, and since the main goal behind smart environments is to assist the user while performing his tasks and to give him a nice experience about the environment, so a successful modeling method should stress on the good understanding of the user tasks in order to be able to deliver a high level of assistance. Several notations for task modeling exist, however the most common one is CTT [9]. It provides a set of temporal operators that express the precise order in which the tasks have to be performed in order to achieve the desired goal.

From this background one can infer the need for patterns being capable to model solutions for any user-related accessibility problem and furthermore these patterns have to be task-based in order to fit into the HCI application domain. In the next section we present our task-based patterns to deal with the user-related accessibility problems in ubiquitous computing and smart environment areas. The notation used for our patterns is the CTT notation.

3 User-Oriented Accessibility Patterns

As already mentioned, in smart environments by using the term 'Accessibility' we mean the capability of any user to receive/deliver any kind of information from/to any existing entity in the environment without being restricted by the information format or any kind of user impairment. In our point of view, any user-related accessibility problem can be categorized as one of the two following types:

1. ***Input accessibility problem***: The user is not able to get information from one or some entities in the environment.
2. ***Output accessibility problem***: The user is not able to express himself and to deliver information to one or some entities in the environment.

Let us take the example of a presentation scenario in a smart meeting room. In order to be able to follow a presentation the user has to be able to listen to the presenter and to read the slides shown on the canvas. Then, let us assume we have a blind user in this meeting room trying to follow the presentation. This blind user is able to get information from the talker while it is impossible for him to read the slides, and so he is having a serious problem to get information from the slides (The information presented on the slides are considered as input for the user).We can describe this situation by telling that our blind user is having an input accessibility problem.

Now, what if the same user is trying to take notes while listening to the presentation? Actually to take notes, one has to write some points either by normal handwriting or using one of his devices (e.g. laptop). Evidently, our blind user is also unable to take notes in the normal and usual way, but this time he fails to deliver information to one of the environment entities and consequently such a problem is categorized as an output accessibility problem, as the environment is unable to receive information from this user.

If we look to the situation from the modeling point of view, how does the problem look like for a developer who is trying to take the accessibility needs into account while constructing his model for a smart environment application? The problem is that there are some tasks which are modeled as atomic actions in the case of normal users, but now these tasks have to be extended in the case of impaired users. So the developer has to integrate an accessibility mechanism or model for each task the user is unable to perform in the usual manner. To overcome this problem and in order to assist the developer to integrate the accessibility needs into his model we present our idea which is composed of two parts. The first step is to construct the input and output generic accessibility patterns, while the second step is to provide two pattern libraries with some concrete instantiations of the generic patterns for the most common accessibility problems. In the following these two ideas are discussed in further details.

3.1 The Input and Output Accessibility Patterns

Any user-related accessibility problem we can imagine is based on the fact that the current type of the information is not suitable for this user, or in other terms our specific user is unable to receive the information in its actual format. That is why our approach relies on the fact that a typical accessibility problem can be solved by providing a mechanism changing the current format of the information to another suitable and understandable format to the user. On a concrete level, this mechanism is a software performing the information's transformation process from the former to the recent format. Such softwares already exist, as an example to change the information format from text to speech the program "*verbose*" [10] can be used, while "*Dragon Naturally Speaking 11*"[11] can be used for the inverse process. The basic idea behind the input and output accessibility patterns is to provide a core model which is able to describe the information format adaptation mechanism regardless of the type of information or the kind of user impairment, and consequently the developer can take any kind of user impairment or accessibility problem into account by just instantiating these patterns and integrating them into his model. The existence of users having some kind of disability can take place in several types of smart environments. For example the probability of having impaired users or elderly people in smart homes is relatively high, as smart homes are able to offer them real assistance while performing their daily life tasks. Consequently, it is crucial that those people be taken into account by the developer while constructing his model for such environments. Moreover smart offices and smart meeting rooms are places where the existence of impaired users is also probable. Therefore we present here our patterns which help the developer to consider these accessibility problems within his model.

As a result, in Fig. 2 we have the generic input accessibility pattern containing the input accessibility mechanism which behaves like a black box having as input the information received by the environment in the usual format and as output the information in the new convenient format processed and understood by the user.

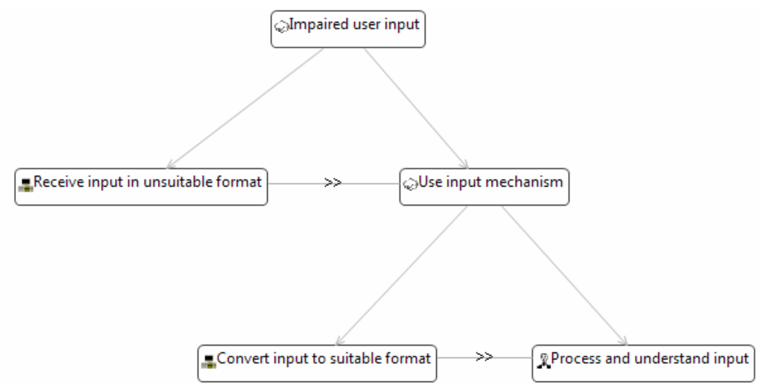


Fig. 2. User-oriented input accessibility pattern

Also in Fig.3 we have the generic output accessibility pattern containing the output accessibility mechanism having as input the information received by the impaired user in the format he can provide it with, and as output the information broadcasted to the environment in the normal format.

3.2 User Impairment-Based Accessibility Patterns Libraries

There is a huge set of possible user impairment types. However, in our point of view the two already mentioned generic patterns can be employed to model any accessibility problem after being instantiated and adapted to the context of use which

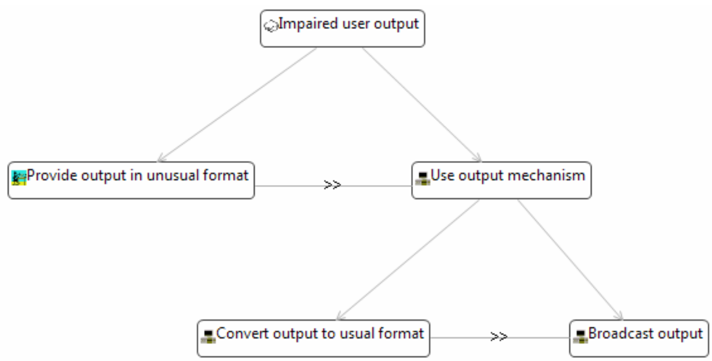


Fig. 3. User-oriented output accessibility pattern

depends on the modality or format of the information and the type of user handicap. In order to maximize the benefit of our patterns we provide here two pattern libraries which contain concrete instantiations out of these generic patterns for the most common and known user impairment types (e.g. blind user, deaf user, paralyzed user…). By having this set of patterns the level of assistance offered to the developer is increased, as for the most expected situations he is able to directly load the convenient pattern from the pattern library without going through any adaptation process. The first library is responsible for any input accessibility problem concerning one of the listed impairment types, and the second library is responsible for the corresponding output accessibility problem. As an example of such patterns we present here the detailed structure of the blind input accessibility pattern. In Fig.4 the solution proposed by the pattern is depicted. Some blind users use screen readers for which the functionality is also based on the idea of changing the modality of the information from text to speech or Braille formats, and so the usage of these screen readers is modeled in the same manner.

Name: Blind input accessibility
Context: A blind user fails to receive information from the environment because this information is presented in unsuitable format (text format).The information type must be changed to a new suitable format in order to be processed and understood by the user.
Problem: How to integrate the information adaptation process into the model.
Solution: Receive the information in its usual format. Use a mechanism (software) in order to convert the information type from the text to the speech format.

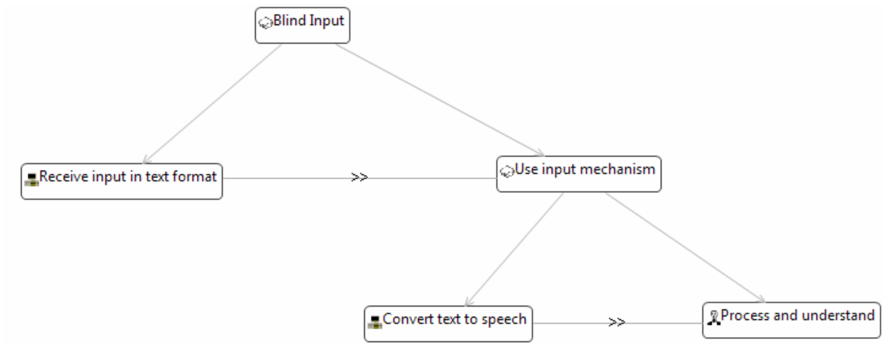


Fig. 4. Blind input accessibility pattern

4 Patterns Application Example

In this section we show an example illustrating the use of our patterns in smart environment application modeling process. In [12] the suitability of CTML (Collaborative Task Modeling Language) as a modeling language for smart environments applications is discussed. Briefly, CTML defines five main steps the developer has to follow in order to model any scenario in smart environment. The first step is to define the roles which are associated to the existing actors in the environment. Let us take the scenario of giving a presentation in a smart meeting room as an example. In this scenario we have three possible roles as we have the chairman, the presenter and the listener. For the sake of brevity, we just focus here on the role 'listener'. In the normal case the role 'listener' is a very simple role to model. A listener has to listen to the talk, take notes and he may ask questions at the end. In Fig.5 this simple case is presented. Normally the listener takes some notes every while during the presentation, and that is why the "Suspend/Resume" temporal operator was chosen before the task 'Takes notes'. Also at the end of the talk this listener may or may not ask questions, and so this task is not mandatory for the accomplishment of his role and consequently the "Choice" temporal operator comes before the task 'Asks questions'.

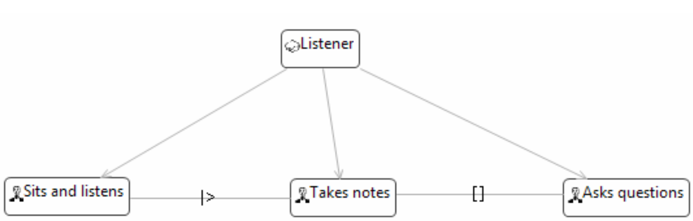


Fig. 5. 'Listener' role for normal user

Now let us think about the same role but in the case of a deaf user. How can the developer use our patterns in order to extend this model and adapt it to the deaf user needs? To have a valid transformation, the developer has to iterate over all the existing tasks and has to substantiate every atomic action the impaired user cannot perform by one of our patterns. To be able to choose the right pattern, the developer has to decide whether the problem for the current task is an input or an output one. So in this example, we will begin by the 'Sits and listens' task. Listening to a presentation implies getting information from the slides shown on the canvas and meanwhile getting information from the presenter. Our deaf user does not have any problem to read the slides but he is incapable of listening to the presenter and consequently it is impossible to get input from the talker. This is an input accessibility problem and so here we need the input accessibility pattern. If we come to the next task 'Takes notes' we figure out that our impaired user does not have any problem to take notes either by handwriting or even using his notebook. So this task does not need to be extended and it remains just the same. Finally when we come to the last task 'Asks questions', most of the deaf users are not able to speak as well. So our user is not able to raise his hand and just begin asking his questions in the usual way, and here we say that the user is having troubles to express his ideas and to deliver information to the environment and so it's an output accessibility problem and the solution is provided by the output accessibility pattern. In Fig.6 the corresponding role 'Listener' for a deaf user is depicted. In fact the developer is able to load the mentioned patterns from the pattern libraries directly as the deaf user case is an expected one. For other kinds of impairment which are not so common he has to use the generic input and output accessibility patterns by instantiating and adapting them to his context of use.

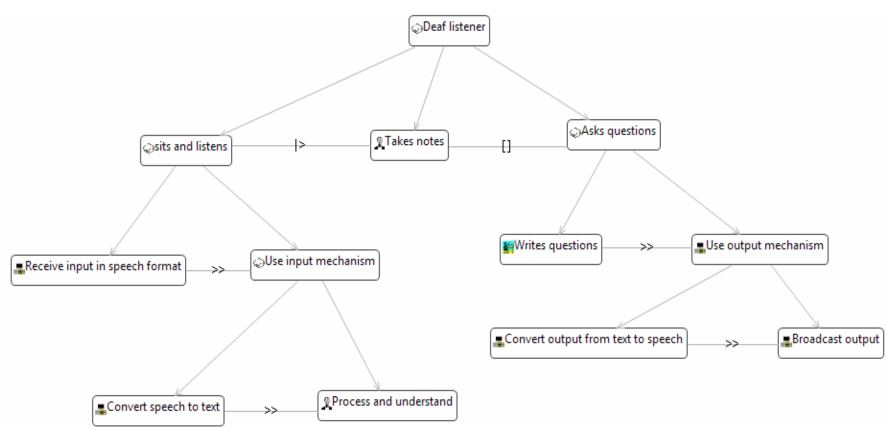


Fig. 6. 'Listener' role for deaf user

5 Conclusion and Future Research

In this paper we discussed the user-related accessibility problems from a different point of view, as we provided a new definition of what accessibility means in the context of smart environments. Based on this definition bidirectional accessibility should be guaranteed to the user, and so we tried to discuss this requirement from modeling point of view. We illustrated the problem as it is seen by the developer by arguing that some atomic actions in the normal user case have to be extended in the case of an impaired one in order to model his accessibility needs. We provided our input and output generic accessibility patterns which can be employed as a core model to be instantiated and adapted to any user-based accessibility modeling situation. Furthermore, to offer a higher level of assistance we suggested two pattern libraries containing instantiations of the two main generic patterns for the most recurrent user-impairment cases. These pattern libraries save the effort and the time consumed by the developer in order to adapt the generic patterns into his context. To illustrate the use of our patterns we presented a simple case study which implements the way these patterns can help the developer in the transformation of his model from the normal to the impaired user case. Finally, we would like to mention that although the scenario tackled in this paper was always the presentation scenario inside a smart meeting room, but the presented patterns can also offer the same level of assistance for modeling applications in other smart environment domains and scenarios as in smart homes for example. Actually the existence of impaired users can even be more common in smart homes, and so the intervention of our patterns can have much more influence in the corresponding scenarios.

In the future we would like to answer the question whether it is possible or not to provide device-oriented accessibility patterns as well and consequently we can build a pattern language addressing any kind of accessibility problem in the whole environment.

Acknowledgments. This work is supported by a grant of the German National Research Foundation (DFG), Graduate School 1424, Multimodal Smart Appliance Ensembles for Mobile Applications (MuSAMA).

References

1. Wurdel, M.: Making Task Modeling Suitable for Smart Environment. In: ICUMT, pp. 1–6 (2009)
2. Alexander, C., Ishikawa, S., Silverstien, M.: A Pattern Language. In: Towns, Buildings, Construction. Oxford University Press, Oxford (1977)
3. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software (1994)
4. Borchers, J.: A pattern approach to interaction design. In: DIS 2000 (2001)
5. Tidwell, J.: Interaction Design Patterns: Twelve Theses. In: PLoP 1998 (1998)
6. Jeschke, S., Pfeiffer, O., Vieritz, H.: Using web accessibility patterns for web application development. In: SAC 2009 (2009)
7. Vieritz, H., Schilberg, D., Jeschke, S.: Merging web Accessibility and Usability by patterns. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6179, pp. 336–342. Springer, Heidelberg (2010)
8. Wurdel, M., Propp, S., Forbrig, P.: HCI Task Models and Smart Environments (2008)
9. Paterno, F., Mancini, C., Meniconi, S.: ConcurTaskTrees: A diagrammatic Notation for Specifying Task Models. In: IFIP TC13, pp. 362–369 (1997)
10. `http://www.nch.com.au/verbose/index.html`
11. `http://www.pcworld.com/reviews/product/587538/review/dragon_naturallyspeaking_11_professional.html`
12. Wurdel, M., Sinnig, D., Forbrig, P.: CTML: Domain and Task Modeling for Collaborative Environments. J. UCS 14, 3188–3201 (2008)

Part IV

Cognitive, Psychological and Behavioural Issues in HCI

Visual Hierarchy and Viewing Behavior: An Eye Tracking Study

Soussan Djamasbi[1], Marisa Siegel[2], and Tom Tullis[2]

[1] Worcester Polytechnic Institute, School of Business, 100 Institute Road, Worcester, MA 01609, USA
djamasbi@wpi.edu
[2] Fidelity Investments, User Experience Group
82 Devonshire Street, V3B, Boston, MA 02109, USA

Abstract. Empirical evidence suggests that users often exhibit a viewing pattern that favors the top and left sides of web pages. This viewing pattern may cause users to miss a great deal of information. Grounded in the model of visual hierarchy, this study examines the impact of visual complexity on how users view a page. The results show that users' viewing pattern in our study was more scattered than those reported in previous studies, which used pages with a relatively less complex visual hierarchy. We also examined the impact of browsing and information retrieval on viewing pattern. Such an investigation can provide insight for the design of homepages that can effectively serve both those who browse and those who retrieve information. The results also show that eye tracker can serve as a valuable tool for designers to develop and test new designs.

Keywords: Eye Tracking, Design Science, Visual Hierarchy, Viewing Pattern, Fixations, Home Pages.

1 Introduction

The Internet has become a valuable source of information for users and a valuable means of communication for companies [1, 2]. Not surprisingly, providing a positive user experience has become a major focus of industry research [3-6]. One way to improve users' web experience is by enhancing the effective communication of key information on a page. Understanding users' viewing pattern can play an essential role in achieving this goal.

According to Nielsen/Norman research group (http://www.nngroup.com/), people tend to exhibit F shaped reading patterns when viewing a page [7]. The F shaped viewing pattern implies that users are likely to miss key information that is not placed on the left or top of a text based page [7-9]. Users missing information on the bottom of a page is particularly evident in studies showing that users focus more on information above the fold, or the part of page that is visible to users without scrolling [10, 11].

Despite the important implications of the F shaped viewing pattern, little work has been done to investigate factors that can affect such a pattern of viewing. Studies have shown that users consistently exhibit an F-pattern of viewing on text-based web

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 331–340, 2011.
© Springer-Verlag Berlin Heidelberg 2011

pages, regardless of whether they are browsing or completing tasks, or whether the text is columnar [12, 13]. However, on purely image-based pages, users do not exhibit the F-pattern [13]. Web pages, and particularly homepages, commonly have a mixture of text and images; viewing pattern on homepages is also of particular importance to firms because research shows that a user's impression of a company can be significantly influenced by the design of the company's homepage [14]. Therefore, this current study expands on previous research by investigating two factors that may affect the F shaped viewing pattern: 1) visual hierarchy, or the arrangement of elements, and 2) task. To conduct these investigations, four prototypes of the homepage from a real company's website were used. Each prototype was manipulated to have a different visual hierarchy. We investigated the impact of two types of tasks, browsing and searching for specific information, on viewing behavior. Because of their relevance to homepages these two types of tasks are often used in usability research [e.g., 10, 13].

2 Visual Hierarchy

According to the model of visual hierarchy, viewing pattern is guided by two distinct cognitive processes: searching and scanning. The first phase, searching, refers to a viewer's attempt to find a point of entry into the page. The second phase refers to a viewer's behavior after finding such an entry point. In this second phase, the viewer extracts information which is located around the entry point [15]. This model suggests that both processes can be influenced by the attributes of the web components. For example, the search process can be influenced by attributes such as the size, color, location, text style and visual information (images) of components. Similarly, the scan process can be influenced by attributes such as proximity and order of components. When used effectively, these attributes create a visual hierarchy that can guide users in viewing a page. For example, by manipulating the size or location of an item, designers can manipulate the order in which web components are viewed and thus influence the search process. Because larger items draw more attention than smaller items, larger objects on a page will be viewed before smaller ones. Similarly, because people exhibit a top down viewing preference, items located at the top of a page will have priority in the visual hierarchy over other items on the page. Designers can also influence the scanning phase of viewing. Because items in close proximity are perceived as related to each other [16], placing related information around an entry point on a web page can facilitate a more effective scan phase.

3 F Shaped Pattern of Viewing

A number of studies show that users exhibit a viewing pattern that is shaped like the letter "F" [7, 8, 17]. An example of such an F shaped pattern of viewing is displayed in Fig. 1. This viewing pattern tends to contain two long horizontal scans and one long vertical scan. The first horizontal scan tends to be longer than the second one [13, 18].

Fig. 1. F-shape pattern of viewing in a text based web page for browsing (the heat map on the left) as well as searching text (the heat map on the right) [13]

According to the above mentioned studies, the F shaped viewing pattern is not affected by task type, such as browsing or searching (e.g., see Fig. 1). A recent investigation, however, shows that, when using web pages composed primarily of images, users do not show an F shaped pattern of viewing. For example, as shown in Figure 2, the top three rows of images receive similar intense fixations while the bottom two rows receive relatively fewer fixations. Moreover, as shown in Figure 2, the fixation pattern for such a page is not independent of the task. Because images often serve as entry points to web pages [15], these findings suggest that visual hierarchy manipulated by images is likely to have a significant impact on users' pattern of viewing. Moreover, such a visual hierarchy may also affect whether users' viewing pattern is task dependent. The next section contains an explanation of how these possibilities were examined though an eye tracking experiment.

Fig. 2. Viewing pattern for an imaged based webpage browsing (the heat map on the left) as well as searching images (the heat map on the right) [13]

4 Eye Tracking Experiment

The theory of visual hierarchy [15] argues that by using size, color, and images, designers can create a visual hierarchy that guides users' attention through a webpage. Such a hierarchy can help users locate entry points and, subsequently, scan for information around those entry points. In other words, according to the theory of visual hierarchy, the design of the page is likely to have an impact on users' viewing pattern. In particular, the visual hierarchy of a page is likely to influence users' tendency to view it in an F-shaped pattern. This viewpoint is supported by the findings of previous research [13]. In particular, results suggest that manipulating the visual hierarchy of a page via images may affect users' F shaped viewing pattern [13].

To test the impact of visual hierarchy on F shape viewing behavior, we used two prototypes of a web page of a real company. These two prototypes were designed to differ in terms of visual hierarchy. One page was designed to have clearer visual hierarchy, with differing font sizes and sectioning (Page A in Fig. 3), while the other was designed to have less clear visual hierarchy, with uniformly sized text (Page B, Fig. 3). The visual hierarchy of each of the two prototypes was further manipulated through the inclusion or exclusion of images of faces. Because images tend to attract users' attention, their inclusion affects the visual hierarchy of the homepage [15]. In this study, images of faces were used because human faces are particularly effective in drawing users' attention [19], and thus can serve as effective points of entry for directing users' attention to information located around them. Based on this, we expected that the inclusion of images of faces on the prototypes to have an impact on users' viewing pattern. In order to examine this, the images of faces were placed in two sections of the page that are typically missed when users exhibit an F shaped viewing pattern: below the fold and on the right side of the page. In the prototype with a clearer visual hierarchy, the faces were placed below the fold of the page, and in the prototype with a less clear visual hierarchy, faces were placed on the right side of the page. The inclusion/exclusion of images of faces on the two prototypes (page A and page B) resulted in a total of four prototypes: two prototypes with clearer visual hierarchy, with and without faces, and two prototypes with less clear visual hierarchy, with and without faces (Fig. 3).

Because task type was of interest in this study, participants were instructed either to browse or to complete an information retrieval task. Because images of faces attract users' attention [5] and serve as points of entry for information that is located in their close proximity [15], their presence on the page may play an important role in users' viewing pattern, both when they browse a page and when they retrieve related information that is placed next to those images.

4.1 Participants and Design

Forty eight (33% male, 67% female) participants in this study were the employees from various industry sectors such as insurance, finance, real estate, and construction. The age of participants ranged from twenty to fifty years. Participants were randomly assigned to one of the two task conditions, browse or information retrieval. Each participant was also randomly assigned to one of the four prototypes (Fig. 3).

4.2 Task

The participants in the task condition were required to retrieve different specific information in random order. For example, the participants were asked to find "an opinion on what's next for GE". Two of these specific pieces of information on the page were located next to images of the faces.

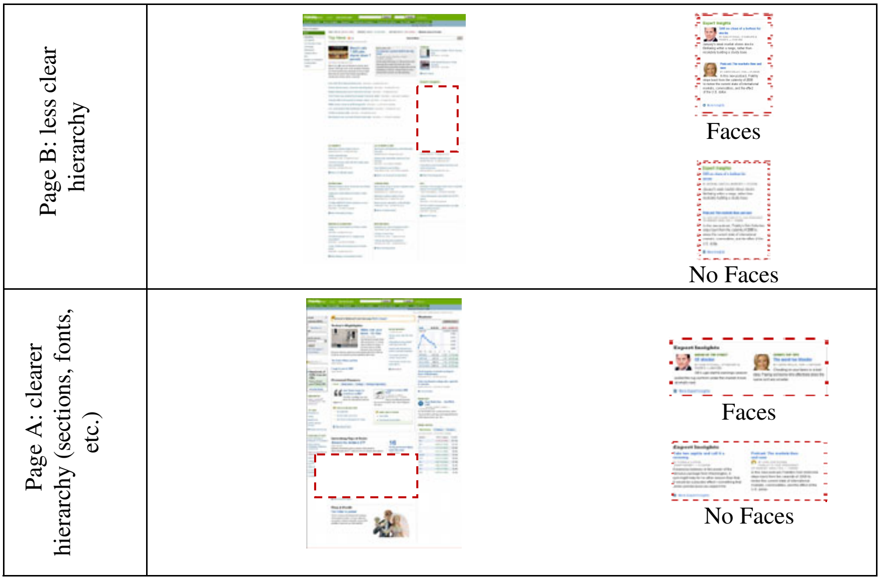


Fig. 3. The prototypes used in our study. The dotted areas show the area where the faces or their textual counterparts were included on the two home pages A and B.

4.3 Measurements

Using the eye tracker, the eye movements and fixations of the participants were tracked. A fixation is defined as a gaze of a minimum of 300 milliseconds and reliably indicates interest [20-22]. Fixation patterns are often analyzed using heat maps, which show how long participants fixated on any given area. Heat maps are typically created using data from several participants, allowing the analysis of fixation trends. Examples of heat maps are shown in Fig. 4. Areas with color were fixated upon, with red indicating longer fixations, and green and yellow representing decreasing lengths of fixation. Areas with no color are those that that did not receive fixations.

4.4 Results

The analysis of the heat maps (Figure 4) indicate that, regardless of experimental condition (task vs. browsing) and prototype (Page A and B with or without faces), users fixated more above the fold of the webpage. This behavior is consistent with the top-down viewing preference suggested by the model of visual hierarchy [15], as well as the F shaped viewing pattern [8, 17]. The fixations in all eight conditions in this study, however, appear more scattered, covering more areas of the page, than the fixations on the heat map from previous studies, shown in Figures 1 and 2. For example, there is much more fixation on the left side of the page, particularly in the information retrieval (search) condition in our study, compared to those shown in Fig. 1 and 2.

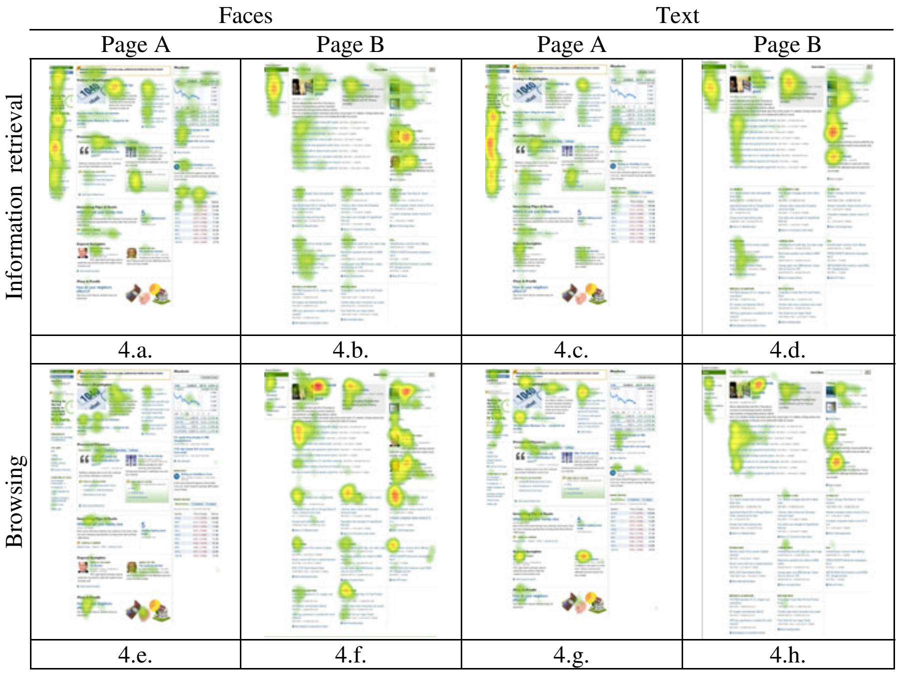


Fig. 4. Heat maps

Comparison of the prototypes provides interesting insight as well. The heat maps for prototypes that did not include faces (4c, 4d, 4g, and 4h) show that, regardless of task condition, Page A received more fixation than Page B. This suggests that the clearer visual hierarchy of Page A retained more attention. Additionally, regardless of task condition, Page A received more fixation below the fold. This is particularly important because the area below the fold often receives little attention from users.

The differences between the pages that did not include faces were not evident in the heat maps for the pages that included faces (4a, 4b,4e, and 4f). In the face condition, Page A contained images of faces below the fold, while in Page B the images were placed above the fold. This suggests that changing the visual hierarchy of Page B by placing faces above the fold had a major impact in attracting users' attention. This is particularly evident in Fig. 4f. The fixation pattern of users in Figure 4f shows that the presence of faces above the fold enticed users to browse the page more meticulously, even below the fold.

The comparison of the task conditions also provides interesting insight. In the browsing condition, the fixations appear more in the central part of the page, as opposed to the left or right sections. This is particularly evident in the Page A prototypes that have clearer left, center, and right sections. Consistent with the principle of visual hierarchy [15], the bright yellow on the text around the image on top left corner of the center section suggests that viewers used this image as an entry point when viewing that part of the page. In the information retrieval condition, in contrast to the browsing condition, there was a high amount of fixation on the left side of page, where navigational links are located (for example, compare Fig. 4a and 4e). There was also a higher amount of fixation on the right side of the page in the information retrieval condition than in the browsing condition. In other words, the information retrieval condition has a more scattered viewing pattern than the browsing condition. These heat maps together suggest that placing information that is important in retaining new visitors in the center of the page is likely to help attract those who browse a homepage. Placing useful links on the left side of a page is likely to be particularly helpful to those who retrieve information on a page.

In the browsing condition, there was a difference in fixation between the prototypes that included faces and those that did not include faces. This was different, however, depending on the location of the faces. Whether participants were browsing or retrieving information, they spent more time fixating on the manipulated section of the page (the Expert Insight section). When faces were above the fold (4f) they attracted longer fixations on the Expert Insight area than when they were not present (4h). The fixations on the prototype with faces (4f) covered more text around the image compared to the prototype without faces (4h), which shows intense fixation only on the title. When faces were placed below the fold, however, they did not seem to attract a great deal of attention. For example, the heat maps in images 4e and 4g show that the Expert Insights section received more intense fixation when faces were not present. Additionally, it is interesting to pay attention to fixations on faces between the two task conditions. When users were browsing rather than were searching for information, faces received more intense fixations. For example, there is more intense fixation on faces in Fig. 4f, as evidenced by a strong shade of yellow on the faces, compared to Fig. 4b, which has green on faces.

5 Discussion

The results showed that the pattern of viewing in this study was different from the pattern shown in a previous study (Fig. 1 and 2). We used prototypes of a homepage that had a more complex visual hierarchy than the previous study. For example, the text in our prototypes utilizes a greater number of attributes (such as size, style, and color) than the text used in Figure 1. The web page used in our study also had fewer images than the page in Figure 2. The combination of text and images, as well as the use of size, color, and location in the main section of our web page creates a more complex hierarchy than the main section of the page in Fig. 2, which contains five rows of equal size images. Thus, the discrepancy between our heat maps and those from previous study indicates that the complexity of visual hierarchy on a page can affect the viewing pattern and guide users to view the page more carefully.

The comparison between the heat maps for the two task types (browsing, information retrieval) in this study showed that the participants concentrated on the central part of the page when browsing, evidenced by shorter fixation on the left or right side of the page compared to fixation on the middle section of the page. When retrieving information, however, the left side of the page (where the links were located) was heavily fixated upon. Moreover, there was higher fixation on the right side of the page (compared to the browsing condition) as well. These results suggest that viewing patterns on the prototypes used in this study were task dependent.

Because the page in this study had a more complex hierarchy than that of the page used in Fig. 1 and 2, the results of this study indicate that visual complexity may affect the effect of task on viewing pattern.

The comparison between the prototypes also showed that the inclusion of images had an impact on viewing pattern. The results show that the images of faces that were placed above the fold attracted the attention of those who browsed the page to text around them. These images also encouraged users to view the page, even below the fold, more carefully. On the prototype without images of faces, viewers fixated more intensely on the headings as opposed to the text beneath. When the faces were placed below the fold they did not attract as much attention as when they were placed above the fold. For those who were searching for information that was located next to the images of faces below the fold, the fixations were less intense on titles when images of faces were available on the prototype compared to when they were not included on the prototype.

These results have important theoretical and practical implications. Studies show that the visual appeal of a page has a significant impact on users' perception of usability [23, 24]. Visual aesthetics is also shown to be a strong predictor of users' willingness to revisit a page [25, 26], suggesting that this work has practical implications for improving the likelihood that website visitors become customers. Because our results show that people fixate more on the center of the page, they suggest that creating visual appeal in this area may be of great importance in creating a favorable aesthetic experience for a user who is browsing a page. These results also contribute to literature by showing that task has an effect on viewing a page with a more complex visual hierarchy. While the participants assigned to browsing concentrated more on the center, those who were tasked with retrieving information focused heavily on the left side of a page. Moreover, this finding is different from previous studies that report a task independent behavior for text-based pages (Fig. 1). More concentration on the left side is also different from the findings of previous studies that show little fixations on the left section of text or image based pages with less complex visual hierarchy (Fig. 1 and 2).

The results indicate that images of faces can serve as an effective entry point to a page. When placed above the fold, these images encouraged users to more carefully inspect the page. When placed below the fold, faces drew the attention of those who browsed the page to the text around them. Additionally, faces attracted stronger fixations when users were browsing the pages. These results suggest that images of faces may play an important role in influencing users' viewing behavior, particularly for those who browse a page. The above results together suggests that eye tracking can serve as an effective tool in developing and evaluating design and thus the analysis of eye tracking data can contribute to human computer interaction research as well as design science.

6 Limitations and Future Research

As with any study, the results of this work are limited to its setting. To minimize the effects of the setting the laboratory was designed to create a natural user setting and the eye tracker used was completely remote from participants, allowing for free movement. Future studies can expand upon this research by including a prototype that replicates web pages in Figure 1 and 2 to more directly compare the effect of visual hierarchy on F-shaped viewing behavior reported in previous studies. Future studies can also extend our results by examining different genres and types of websites. For example, a retail homepage has different goals than a financial homepage and, therefore, is naturally arranged differently. As a result, these two different types of pages can result in different viewing patterns.

7 Conclusion

This research indicates that the F shaped viewing pattern is affected by the visual complexity of a webpage, and that the type of task does have an effect on the way users view a more complex homepage. These results are consistent with the theory of visual hierarchy and show that a more complex visual hierarchy can guide users' attention to areas outside of the F shaped pattern of viewing found on less visually complex pages. These results have both theoretical and practical implications. On a broad level, theoretically, this research contributes to design science and human computer interaction research by providing insight about users' fixation patterns when browsing and when retrieving information from a homepage. More specifically, this research expands prior findings on the effect of task on user behavior, as well as the effect of images and, more specifically, images of faces. Practically, this work provides additional understanding of how users view pages. This allows companies to provide a better user experience, which in turn can help attract more users to their websites.

References

1. Goodwin, B.: Rush to Make the Most of Web 2.0 Sends Developers' Salaries Soaring. Computer Weekly, 48 (2006)
2. Hammer, M., Champy, J.: Reengineering the Corporation: A Manifesto for Business Evolution. Harper Business, New York (1993)
3. Djamasbi, S., Tullis, T., Hsu, J., Mazuera, E., Osberg, K., Bosch, J.: Gender Preferences in Web Design: Usability Testing through Eye Tracking. In: Proceedings of the 13th Americas Conference on Information Systems, Keystone, Colorado, 1-8 (2007)
4. Djamasbi, S., Tullis, T., Girouard, M., Hebner, M., Krol, J., Terranova, M.: Web Accessibility for Visually Impaired Users: Extending the Technology Acceptance Model (TAM). In: Proceedings of the Twelfth Americas Conference on Information Systems, Acapulco, Mexico, pp. 3029–3034 (2006)
5. Djamasbi, S., Tullis, T., Siegel, M., Ng, F., Capozzo, D., Groezinger, R.: Generation Y & Web Design: Usability Testing through Eye Tracking. In: Proceedings of the Fourteenth Americas Conference on Information Systems, Toronto, Canada, pp. 1–11 (2008)
6. Tullis, T., Siegel, M., Sun, E.: Are people drawn to faces on webpages? In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4207–4212 (2009)
7. Nielsen, J.: F-Shaped Pattern For Reading Web Content (2006), http://www.useit.com/alertbox/reading_pattern.html

8. Shrestha, S., Lenz, K.: Eye Gaze Patterns while Searching vs. Browsing a Website. Usability News 9 (2007)
9. Shrestha, S., Owens, J.W.: Eye Movement Patterns on Single and Dual-Column Web Pages. Usability News 10, 1–7 (2008)
10. Buscher, G., Cutrell, E., Morris, M.R.: What Do You See When You're Surfing? Using EyeTracking to Predict Salient Regions of Web Pages. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 21–30 (2009)
11. Bailey, R., Koyani, S., Nall, J.: Usability testing of several health information Web sites Technical Report. National Cancer Institute, 7-8 (2000)
12. Shrestha, S., Owens, J., Chaparro, B.: The Effect of Location and Congruency of Text Ads on Information Search. In: Human Factors and Ergonomics Society Annual Meeting Proceedings, vol. 53, pp. 1262–1266 (2009)
13. Shrestha, S., Lenz, K.: Gaze Patterns while Searching vs. Browsing a Website. Usability News 9, 1–9 (2007)
14. Oh, J., Fioritob, S.S., Cho, H., Hofackerd, C.F.: Effects of Design Factors on Store Image And Expectation of Merchandise Quality in Web-Based Stores. Journal of Retailing and Consumer Services 15, 237–249 (2008)
15. Faraday, P.: Visually Critiquing Web Pages. In: Proceedings of the 6th Conference on Human Factors and the Web, Austin, Texas, pp. 1–13 (2000)
16. Flieder, K., Mödritscher, F.: Foundations of a pattern language based on Gestalt principles. In: CHI 2006, Extended Abstracts on Human Factors in Computing Systems, pp. 773–778 (2006)
17. Wilkinson, S., Payne, S.: Eye tracking to identify strategies used by readers seeking information from on-line texts. In: Proceedings of the 13th Eurpoean Conference on Cognitive Ergonomics: Trust and Control in Complex Socio-Technical Systems, pp. 115–116 (2006)
18. Nielsen, J.: F-Shaped Pattern for Reading Web Content (2006), `http://www.useit.com`
19. Tullis, T., Siegel, M., Sun, E.: Are people drawn to faces on web pages? In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4207–4212 (2009)
20. Pan, B., Hembrooke, H., Gay, G., Granka, L., Feusner, M., Newman, J.: The Determinants of Web Page Viewing Behavior: An Eye Tracking Study. In: Proceedings of the 2004 Symposium on Eye Tracking Research & Applications, pp. 147–154 (2004)
21. Chapman, P.: Remembering what we've seen: Predicting recollective experiences from eye movements when viewing everyday scenes. In: Underwood, G. (ed.) Cognitive Processes in Eye Guidance, pp. 237–258. Oxford University Press, UK (2005)
22. Rayner, K., Liversedge, S.P., White, S.J., Vergilino-Perez, D.: Reading disappearing text: cognitive control of eye movements. Psychological Science 14, 385–388 (2003)
23. Lavie, T., Tractinsky, N.: Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. International Journal of Human-Computer Studies 60, 269–298 (2004)
24. Lindgaard, G., Dudek, C.: User satisfaction, aesthetics and usability: Beyond reductionism. In: Proceedings of IFIP 17th World Computer Congress, Montreal, Canada, pp. 231–246 (2002)
25. Loiacono, E.T., Watson, R.T., Goodhue, D.L.: WebQual: An instrument for consumer evaluation of web sites. International Journal of Electronic Commerce 11, 51–87 (2007)
26. Loiacono, E.T., Watson, R.T., Goodhue, D.L.: WebQual: a measure of Web site quality Marketing Theory and Applications (2002)

Cognitive Analysis for Knowledge Modeling in Air Traffic Control Work

Satoru Inoue[1], Hisae Aoyama[1], and Keiichi Nakata[2]

[1] ATM Department, Electronic Navigation Research Institute,
7-42-23, Jindaiji-higashi-machi, Chofu, Tokyo 182-0012 Japan
[2] Informatics Research Centre, University of Reading,
Whiteknights, Reading RG6 6UD United Kingdom
{s-inoue,aoyama}@enri.go.jp, k.nakata@henley.reading.ac.uk

Abstract. Air Traffic Control systems are a kind of service which allows controllers from the ground to keep aircraft safely separated to avoid collisions. It is important to use cognitive systems for the development of this type of critical system. Systems should not only have high performance functions, but also have better user friendly functions and accessibility. It can be said that we have to analyze the functional elements in the work of systems design. To design more reliable interfaces or training programs for the provision of ATC systems, we need to understand the details of basic functions of air traffic controller's tasks within the system. In this research, we discuss the analysis of ATC tasks and modeling of the knowledge of air traffic controllers. And then, we attempt to formalize the basic knowledge of controllers to help them to have a good understanding of the knowledge structure and logical relations.

1 Introduction

Air Traffic Control (ATC) services need to be smoother and more efficient in order to respond to increasing air traffic demands. In current ATC systems, air traffic controllers control air traffic flow by way of their decision making using computer tools.

Air traffic controllers are expected to provide services to keep the air space safe and maintain smooth air traffic flow On the other hand, as the work and tasks of controllers become more complex and the volume and types of information required to carry out these tasks becomes increasingly greater and more complex, the need for systems that are designed to support controllers becomes greater (Banks, 2002). Thus, supporting systems are necessary to enable appropriate cooperation between air traffic controllers and systems in the performance of ATC work in the near future.

With human-machine systems, it is important that design consideration is given to each role and relationship between the human and the machine in order to achieve appropriate cooperation. This is also a key issue for design in ATC systems. On the other hand, however, it is difficult for designers and developers to understand the knowledge content of air traffic controller work, with that knowledge being highly specialized.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 341–350, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Besides acquiring basic knowledge, air traffic controllers undergo simulator training to gain sufficient experience for the acquisition of specialized knowledge and skills. The knowledge needed for ATC operation is not sufficiently formalized, as a large part of ATC operation depends on the experience of each Air Traffic Controller.

We see a need to analyze and model structures of objective work to understand this tacit knowledge and skills. In this research, we first discuss the task analysis method based on distributed cognition for analyzing the real situation of ATC. After that, we also discuss the framework of knowledge management for Air Traffic Controllers. In this research, we attempt to formalize basic controller knowledge to facilitate the controllers' understanding of knowledge structures and logical relations. The aim of this project is to study the applicability of distributed cognition to such knowledge management tasks, with the objective of developing a systematic framework to represent knowledge relevant to ATC expertise for support.

2 Cognitive Process Perspective in ATC Systems

Future air traffic control systems will become more complex and more accomplished by including the element of management. The cognitive systems approach plays an important role in the development of such complex systems. However, even if s systems is complex, they do not need to have many high performance functions, but should always be user oriented and provide greater accessibility. So, there is a need to analyze the functional elements for designing systems. In order to design a system that can assure system safety, enhanced usability, and increased operator reliability, it is critical for the developers of ATC systems to consider the specifics of how the control system is operated, as well as the cognitive characteristics of controllers.

Automation tools have been used as effective support tools in various industrial sectors. However, human error occurs when the mutually dependent relationships between controllers and machines break down. A promising strategy for systems to assist in task performance is the concept of cognitive systems that try to enable human-systems interactions in a knowing manner that is similar to the way in which humans interact with one another (Forsythe et al, 2006).

Such systems require a user model that explains user behaviour from various aspects of cognitive processes such as awareness, memory, user knowledge and experience, context recognition, planning, intention formation, and even consciousness. User models can be used to predict a users' cognitive processes which in turn can be used to better support them (Haikonen, 2003).

In particular, from the cognitive process perspective, it is essential that systems developers understand the complex working processes that are involved in cooperative work by multiple controllers.

However, little has been studied about the cooperative work of air traffic controllers using analytical methods. Since air traffic controller skills are acquired through specialized training, their cooperative work processes are very complex and temporal constraints are also very restrictive and severe. In order to design and develop more reliable systems or training programs for controllers for the future ATC systems, we need to understand the details of the basic system (including controller) functions.

Distributed cognition is a methodological framework by which to analyze the cognitive processes that span multiple actors mediated by technology (Hollan et al, 2000). Distributed cognition can be effective in analyzing cooperative work from the cognitive process perspective. A central tenet of distributed cognition is that cognition should be regarded as a property of a system of individuals and external representational artifacts carrying out cooperative activities (Fields, 1998). Distributed cognition analysis makes explicit the dependencies between human actors and artifacts by examining the transformation and propagation of information through the various forms of representations. As such, 'knowledge' can be represented in terms of interactions in context, which lends itself to further analysis. The management of knowledge, and hence the retention of knowledge, is through changes in distributed cognition induced by the introduction of new systems, personnel, and norms.

3 Task Analysis Based on Distributed Cognition

3.1 Distributed Cognition

In this research, we attempt to analyze and model interactions that take place in current en route ATC work based on distributed cognition. We have taken the activities of a cooperative team of en route controllers as our unit of analysis from a cognitive process perspective. We discuss the application of ethnographical analysis in en route controllers' work as a team, and report on the findings from our ethnographical analysis, followed by a description of analytical models.

An ethnographic approach can be effectively applied when the problem involves the analysis of what knowledge and experience people use in the context of cooperative work. Ethnomethodology is a method of sociology for determining the implicit orders, rules, or norms behind human activities through observation in the actual work environment.

In this paper, we focus on the factors of team performance. As a first step, to analyze how air traffic controllers work, we carried out data collection through the observation and recording of actual work activities in the Tokyo Area Control Centre (TACC) control room.

3.2 ATC Work Setting

We observed that there are some specific and characteristic points in ATC work; in particular, the basis of that work is prediction and instruction, to secure and maintain a safe traffic situation.

Air traffic controllers control air traffic separated into many distinct areas called sectors and more than two controllers are allocated to and take charge of each sector as a team. Usually one controller takes the role of radar controller, and the other takes that of coordination controller. An additional controller will join in for a busy sector, but twin controllers as shown in Fig. 1 are assumed for this work.

Fig. 1. Working situation of Japanese en route ATC

We carried out observation during a time period of relatively heavy traffic that imposed a certain level of workload on the air traffic controllers. Two fixed video cameras were used for recording the behaviour of the controllers at work, and another was used for recording the radar screen of an auxiliary control console that displayed the same radar image as the controllers were looking at on their work console. A digital audio recorder was set up above the control console to record conversations between the radar controller and the coordination controller. In addition, radio communication between the radar controller and airplane pilots, and between the coordination controller and different sectors, and traffic data such as the position, ground speed, and altitude of aircraft, which were displayed on the radar screen and stored in the control centre computer , was obtained.

In our case, the target of observation was a sector called “Kanto-north” (T03) shown in Fig. 2, which spreads over the northern area of Tokyo. The radar controller in en route control predicts and estimates traffic from five to ten minutes into the future. Meanwhile the coordination controller elaborates on these instructions to keep

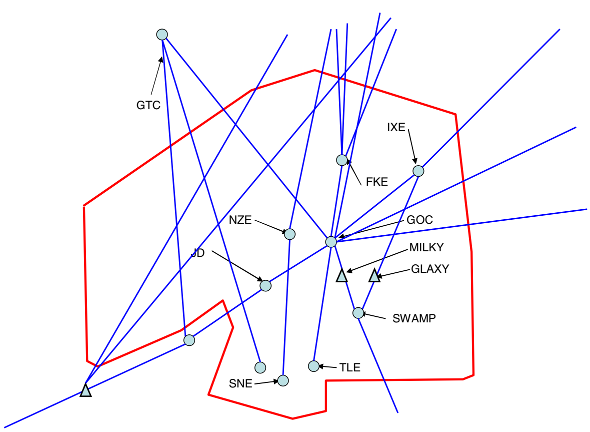


Fig. 2. Kanto-north sector

a safe separation of aircraft based on current information. Many interruptions occur when the controllers have to handle more than two aircraft at the same time: calls from other aircraft outside the immediate focus, requests for handoff from another sector, etc. The coordination controller, who has similar tasks to the coordinating tasks of neighboring sectors, also has to deal with interruptions. In addition, the controllers have to control all IFR aircraft in their own sector. Since en route ATC work has to deal with a variety of states and conditions in their sector, it differs significantly from well formalized tasks such as assembly line operation.

3.3 Data Gathering for Analysis

In this observation, we recorded motions and sounds by video, and system logs as basic data for the analysis. From these we reconstructed the controllers' actions and protocol logs, and analyzed the controllers' tasks in each situation. The system had functions to record multiple types of time-series data such as video, audio and operation logs. We obtained this data for a total of 6 hours by spending 3 days in observation.

Video data
The Video recorded the air traffic controllers' behaviour such as instruction, and coordination, etc., in the control room. Cameras recorded the entire control room from three directions including the radar screen, the flight-data-strip bay, and the view from the back of the room. Moreover, we combined the video from all of the cameras and the audio from the radar controller, the coordination controller, and the pilot, synchronized them, and then recorded them in a batch.

Flight-data-strip
Markings, notes, etc., for flight-data-strips are written by the air traffic controller during their control work. We can find clearance data and instructions for individual airplane pilots and the content of coordination with other sectors in the records on these strips.

3.4 Overview of Data Analysis

In the analysis of the video data, audio data, and radar screen images, they were firstly combined into a single track of video-audio data with synchronized time stamps. Radio communications and conversations among ATCos were then transcribed, and the speakers and listeners of these conversations were identified. The actions of controllers were next recognized from the video data and added to the transcribed protocol data. The data derived from these conversational records and behavioral records, and so on were segmented by the basic units of ATC instructions. With the help of a rated ATCo, we clarified the relationships between the segments and identified the expert knowledge and judgment behind them. This analys showed that Rader Controller and Coordination Controller shared control strategy and team intentions through both verbal and nonverbal communications and established mutual beliefs on SA. These processes lead to smooth cooperation in en-route ATC.

4 Definition and Description of Knowledge

It is not easy to acquire the knowledge needed for ATC work as ATC operations are complex and performed as a team. To codify and manage this knowledge can assist in effective knowledge acquisition. However, explicit knowledge is only part of the whole. Therefore, we need to examine what knowledge is tacit and what knowledge is explicit in ATC work. This chapter will discuss the definitions and representations of knowledge.

4.1 Theory of Knowledge Management

Within the general concept of knowledge management, there is "tacit knowledge" and "explicit knowledge." Tacit knowledge is individual knowledge which is personal and context-specific. But, such knowledge is difficult to formalize and communicate concisely to others.

On the other hand, explicit knowledge can be represented by a formal, logical, systematic language (Polanyi, 1958). Polanyi wrote that "human beings acquire knowledge by actively creating and organizing their own experiences." Given this concept, knowledge that can be represented by systematic language or symbols is part of overall knowledge. Tacit and explicit knowledge are closely related; however, they each have different aspects. Tacit knowledge is subjective while explicit knowledge is objective. Nonaka and Takeuchi (1995) refer to explicit and tacit knowledge, saying that "the dynamic model of knowledge creation is anchored to a critical assumption that human knowledge is created and expanded through social interaction between tacit knowledge and explicit knowledge," and that the process, called "knowledge conversion," is cyclical and mutual. Knowledge conversion forms a process of transformation called the “knowledge spiral,” shown in Fig.3. This cycle is a dynamic process in organizational activity (Nonaka et al. 1994).

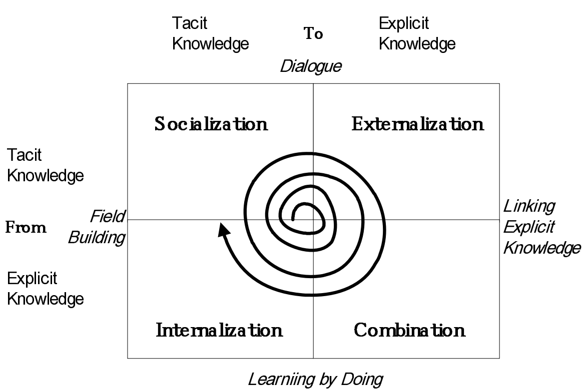


Fig. 3. Four modes of knowledge conversion, and the knowledge spiral (Nonaka & Takeuchi, 1995)

One purpose of codifying a systematic structure of knowledge is to assist the transformation process of these modes of knowledge. In order to formalize the knowledge of air traffic controllers in ATC work, it is important to analyze the relationships between work processes and to understand the specific content of tacit knowledge and explicit knowledge. In the next section, we will discuss knowledge forms in ATC work.

4.2 Composition of Knowledge

If we learn to master technical skills and special knowledge, to demonstrate the good performance of a system that includes humans, there must be a formal, systematic, codified training program in which logical thinking, experience and skills are combined. The knowledge conversion process can thereby proceed smoothly and efficiently by formalizing the theory and experiences described in the previous section. The framework should enable the efficient organization and acquisition of special knowledge and skills by assisting in the process of knowledge conversion, such as the tacit knowledge creation or acquisition of explicit knowledge. Air Traffic Controllers have special knowledge and skills for carrying out ATC operations. We therefore consider knowledge management by Air Traffic controllers to be an important element for the future in ATC systems from the perspectives of safety and training.

In ATC work, air traffic controllers learn about conventional rules and technical regulations which they first encounter in manuals. They then acquire basic knowledge of and skills for control operations through iterative simulator training. After that, controllers learn specific rules, technical regulations, and orders for the sector where they will be in charge of operations. They finally qualify as controllers after acquiring through training the knowledge and skills needed to perform operations in the sectors where they will work. Here, explicit knowledge such as rules and regulations are captured and provided to the controllers as documents, while tacit knowledge is acquired through the training and application of learned knowledge in developing their skills. As such tacit knowledge is more embedded in the context of work and depends on audio-visual perception, situational awareness, and norms that are not easily represented and captured. The following definitions elaborate on explicit knowledge and tacit knowledge as used in this study.

Explicit knowledge

Knowledge that can be written in a systematic language in a document. The most basic ATC operational information is written in the form of regulations and rules. Moreover, knowledge which can be described in a systematic language following analysis and evaluation is assumed to be explicit knowledge, even if the knowledge is based on empirical measurements or techniques involving the exercise of an individual's skills.

Tacit knowledge

By analyzing and arranging the data pertaining to knowledge and skills, there are some types of information more easily understandable as non-linguistic visual or aural stimuli than in a systematically linguistic form. Knowledge or skills of this type, such as situational awareness, temporal awareness, timing of actions, and interaction through space, are defined as tacit knowledge, which is difficult to represent linguistically.

4.3 Representation of Explicit Knowledge

We have examined the work of controllers in terms of distributed cognition, and analyzed task-flow knowledge. We will now discuss the development of a knowledge structuring tool which focuses on the management of explicit knowledge. We report here some results of our attempt to formalize and express explicit knowledge, which forms part of the development of a knowledge management framework.

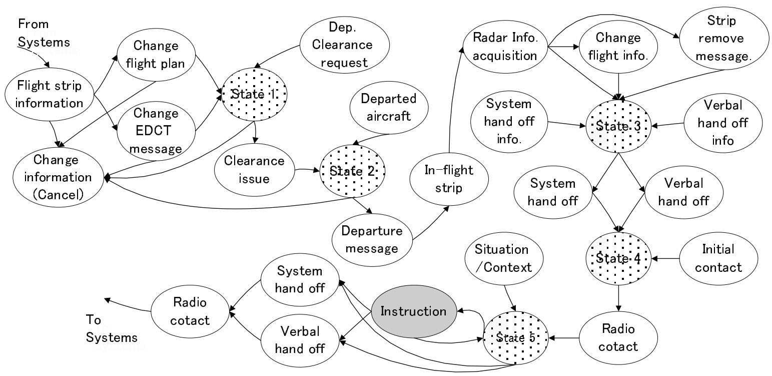


Fig. 4. Concept level of knowledge network diagram in ATC work domain

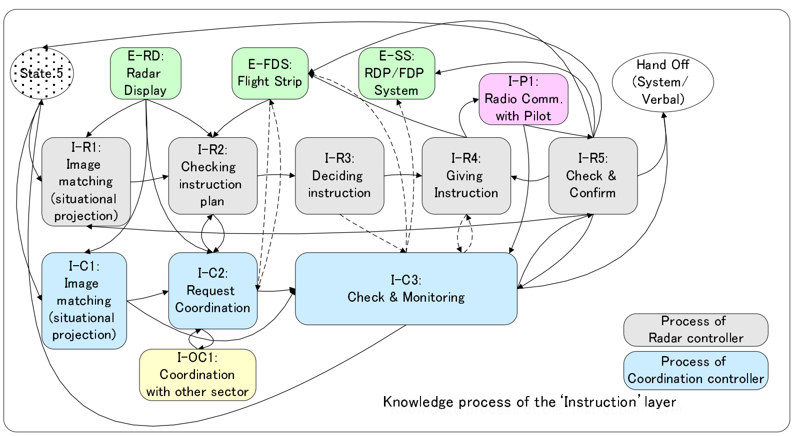


Fig. 5. Knowledge network structure into the "Instruction" layer

The relationships between the forms of knowledge that are applied in ATC work is represented in Fig.4 and Fig.5 as a knowledge network, formulated on the basis of observation and interviews. ATC task processes are expressed as state transitions in the knowledge network. The knowledge network has a layered structure, but how the structure is layered depends on how the knowledge is quantized. In this regard, appropriate quantization techniques are not discussed. The advantage of having a task-flow clarified by distributed cognition analysis expressed as a network of knowledge is in being able to arrange the relationship between the form of work an air that an traffic controller engages in and the information that he/she uses. It is difficult to arrange information regarding the means of use of an entire body of knowledge, because even if each state uses the same information, when expressing it in a task–flow diagram that information has to be described discretely for every state; however, by using a knowledge network, it is easy to understand the state of the information use, as each state is arranged by describing its relationships with other states in relation to that state. Moreover, it is effective in that tacit relations are re-written as explicit knowledge, thus clarifying till then unseen informational relationships. We attempt to describe this knowledge structure using the prototype tool of the knowledge management framework.

ATC knowledge is roughly divided into 2 types: knowledge about rules and regulations, and knowledge related to skills for operation. We tried to make a model of skills knowledge analyzed using distributed cognition: a tool to describe the classifications and knowledge relationships by tagging each state of a node in a knowledge network. In addition, the rules for the task process, which are in turn invoked by the trigger of state transition, can be described as system properties at the node. The conditions and parameter items which describe the system properties can be considered a knowledge model of the decision making process for air traffic controllers. We consider the tagging of system properties to be effective in shedding light on the relationships between, and attributes comparing, other processes. Codifying and managing system property tags as knowledge has promise for the standardization and improvement of training. As a result, we believe that distributed cognition analysis can contribute to improved reliability in controllers and safety in ATC systems.

5 Conclusion

Distributed cognition is applicable to knowledge management tasks where the objective is to develop a systematic framework representing knowledge relevant to ATC expertise. We have proposed a technique for knowledge management as an approach to the study of problems relating to the human factor in an ATC system, and to the structuring of ATC tasks in order to overcome these problems.

The formalized techniques of explicit knowledge have already been covered above; however, the issue of how to represent tacit knowledge in our frame-work of knowledge management remains to be addressed. An approach that describes and represents information in the form of images, a movie or a role-playing simulation (such as a serious game) may be necessary in acquiring knowledge which is difficult to describe or understand in a linguistic form. Learning from a story of a situation can be effective in acquiring tacit knowledge. We are still examining the issue of the representation of tacit knowledge for a future project.

References

1. Banks, S.B.: Perspectives on the State of Modeling and Simulating Human Intent Inferencing. Inference for Users, Teams, and Adversaries, AAAI FS-02-05, 2-10 (2002)
2. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors 37(1), 32–64 (1995)
3. Fields, R.E., Wright, P.C., Marti, P., Palmonari, M.: Air Traffic Control as a Distributed Cognitive System: a study of experimental representations. In: ECCE-9 (1998)
4. Forsythe, C., Bernard, L.M., Goldsmith, E.T.: Cognitive Systems: Human Cognitive Models in Systems Design. Lawrence Erlbaum Associates, Inc., Mahwah (2006)
5. Hikonen, O.P.: The Cognitive Approach to Conscious Machines, Imprint Academic Com (2003)
6. Hollan, J., Hutchins, E., Kirsh, D.: Distributed Cognition. ACM Trans. on Computer-Human Interaction 7(2), 174–196 (2000)
7. Inoue, S., Aoyama, H., Kageyama, K., Furuta, K.: Task Analysis for Safety Assessment in En-Route Air Traffic Control. In: Proceedings of 13th International Symposium on Aviation Psychology, pp. 253–258. OK, Oklahoma (2005)
8. Nonaka, I., Byosiere, P., Borucki, C.C., Konno, N.: Organizational Knowledge Creation Theory: A First Com-prehensive Test. International Business Review 3(4), 337–351 (1994)
9. Nonaka, I., Takeuchi, H.: The Knowledge Creating Com-pany: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, Inc., Oxford (1995)
10. Polanyi, M.: Personal Knowledge. University of Chicago Press, Chicago (1958)

Individual Differences in Work Load While Doing Multitasking with a Computer

Kari Kallinen, Inger Ekman, and Niklas Ravaja

Aalto University School of Economics, Center for Knowledge and Innovation Research
P.O. Box 2125, FIN-00076 Aalto, Finland
{Kari.Kallinen,Inger.Ekman,Niklas.Ravaja}@aalto.fi

Abstract. In the present study we examined the individual differences in work load while doing high and low cognitive load multimodal (audive + tactile or visual + tactile) tasks. We found among other things that participants that are characterized as having high information capacity had lower levels of work load and shorter reactions times during easy and difficult multitask as compared to participant that are characterized as having lower level of information processing capacity.

Keywords: Multitasking, multimodal information, context awareness, cognitive task load, individual differences, tactile, audio.

1 Introduction

Modern work is involving more and more usually multitask technological and information environments. The information load and modalities have been increasing rapidly as the sophisticated computers can now present and take advantages of all human senses: ears, eyes, touch and movement. As people are loaded with increasing amount of information in different modalities, it is important to know the interactions, limitations and individual differences in coping with the multimodal information to be able to build effective information systems and interfaces. For example, an inappropriate information system in critical work environments, such as emergency and military environments, may have serious consequences.

The interest in the research on multitasking and support for task awareness in critical and other information work has rapidly increased in last few decades. The present work is related to a project called "Supporting Situation Awareness in Demanding Operating Environments through Wearable Multimodal Interfaces (SAWUI)". One of the very central aims of the project is to study the limitations of human information processing as well as the possibilities of multimodal information in user interfaces (See Laarni et.al, 2010). In our previous study, we examined the usability of tactual information during high and low cognitive load tasks (see Kallinen and Ekman, 2010). We found for example, that people were able to detect tactual information quite fast and reliable even though their attention was occupied in other task. However, even though there already seems to be a quite large amount of research on multitasking (in addition to our previous study), there seem to be fewer

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 351–358, 2011.
© Springer-Verlag Berlin Heidelberg 2011

studies on the potential individual differences that moderate the ability to perform multiple tasks simultaneously. And yet it is obvious that people differ in their capacity for information processing and coping with mental load for example. These individual factors may play an important role in how people manage in critical information processing situations. Task awareness and multitasking systems should be designed so that they would be optimal for individual users.

In the present study we were especially interested in the BIS and BAS dimensions of personality, information processing capacity, and capacity to cope with stress. We expected that people with high information processing capacity and capacity to cope with stress would experience less work load even in hard work conditions as people with lower levels of such skills. BIS and BAS sensitivities are also potential moderators of task performance, given that BIS regulates aversive motivation and the BAS regulates appetitive motivation. It has been suggested that people characterized as extroverts (e.g, high BAS scorers) are generally under aroused and seek greater stimulation than introverts (e.g. high BIS scorers) to keep their arousal in an optimal level.

2 Method

2.1 Participants

Participants were 34 healthy Finnish men, from 20 to 59 years of age (*M*=30.62). Their average height was 180 cm and weight 79.7 kg. They evaluated that they were in pretty good shape (M=3.71 on a 5 point scale) and that their ability to cope with stress was good (M=3.65 on a 5 point scale).

2.2 Stimuli

Stimuli material consisted of 5 Finnish army voice commands (e.g., Halt! Password!), 5 pictures that presented military equipment (e.g., a tank), and tactual stimuli that were given to four locations in both right and left side of the body: head, neck, arm and torso.

Engineering Acoustics Inc. C2 tactors were used to give the tactual stimuli (see Figure 1).

2.3 Measures

Demographics and BIS/BAS Sensitivies. Background factors such as age, gender, and level of stress management skills as well as Behavioral inhibition (BIS) and Behavioral activation (BAS) sensitivities, were assessed with questionnaires before the experiment. Dispositional BIS and BAS sensitivities of the participants were measured with the BIS/BAS scales (Carver & White, 1994), a 20-item self-administered questionnaire. The BIS scale is comprised of 7 items (e.g., “I feel pretty worried or upset when I think or know somebody is angry at me”). The BAS scale is comprised of three subscales: BAS Drive (DR) reflects the persistent pursuit of desired goals (4 items; e.g., “I go out of my way to get things I want”); BAS Fun Seeking (FS) reflects both a desire for new rewards and willingness to approach a

Fig. 1. EAI C2-tactor

potentially rewarding event (5 items; e.g., "I crave for excitement and new sensations"); and BAS Reward Responsiveness (RR) reflects the orientation to respond positively to the occurrence or anticipation of reward (4 items; e.g., "When I get something I want, I feel excited and energized"). Each of the items was rated on a 4-point scale, ranging from 1 (very false with me) to 4 (very true for me). The psychometric properties of the instrument have been shown to be acceptable (Carver & White, 1994).

Work Load. The NASA TLX (see http://humansystems.arc.nasa.gov/groups/TLX/) was used to assess the work load. The overall workload is a summary of the following subscales in the questionnaire: mental load, physical load, temporal load, performance, effort and frustration. Each scale was rated with a 9 point gradual scale from very low to very high. In addition to overall work load, we also analyzed the mental load subscale separately.

Task Performance. The task performance was measured by the reaction time to the tactual stimuli.

2.4 Procedure

After a brief description of the experiment, participants filled in the background and BIS/BAS questionnaires.

Participants were told that the experiment involve multitasking. They were advised to respond as fast as possible to pictures presented on a screen or sounds presented in earphones. They were told to click a button whenever a picture or a sound they saw/hear was the same as previous picture or sound (in N1, low cognitive load, easy task) or the same as picture or sound as the picture or sound before the previous picture or sound (in N2, high cognitive load, difficult task). They were also told that during this so called n-back task stimuli, tactile stimuli will be presented randomly to different parts of their body and that they have to respond immediately to the tactile stimuli by pushing a button.

There were four approximately 20 minutes blocks of stimuli that were the following: N1 auditive, N2 auditive, N1 visual, N2 visual. The order of the presentation of the blocks was counterbalanced. After each block the subject filled the NASA TLX questionnaire, and had a short break before the next block of stimuli.

2.5 Study Design and Analyses

The study design was Difficulty (2 levels: low, high) x Presentation Modality (2 levels: audio, visual) within-subjects design. The data for the NASA-TLX was analyzed using GLM Repeated Measures Analysis and the data for the reaction time using Linear Mixed Model in SPSS.

3 Results

The analysis confirmed the manipulation for the task difficulty: the N2 task (difficult task) elicited higher NASA-TLX scores and longer reaction times to tactual stimuli than N1 (easy task; For NASA-TLX *M*s = 46.25 and 42.00, $p<.001$; for reaction times *M*s=1.33 and 1.20, $p<.001$).

3.1 Work Load (NASA-TLX)

In predicting work load (NASA-TLX scores), a significant Task Difficulty (N1, N2) x BAS interaction was found (p=.006). As illustrated in Figure 2, among participants scoring high on BAS there was not so large difference in workload ratings between the difficult and easy task (*M*s = 42.52 and 39.97), whereas among low BAS scorers the difficult task elicited much higher work load than the easy task (*M*s = 50.44 and 44.28).

The analysis revealed also a significant Task Modality x BIS interaction in predicting work load (p=.031). The auditive modality elicited lower workload than visual modality among high BIS scorers (*M*s=41.09 and 47.06), whereas among low BIS scorers there was not so large difference (*M*s = 42.93 and 45.42).

The GLM Repeated Measures analysis revealed also a significant Task Difficulty x Information Processing Capacity in predicting workload (p= 0.33). As illustrated in Figure 3, N2 was especially more difficult than N1 among people with low information processing capacity (*M*s = 52.41 and 44.86), whereas among people with high information processing capacity the difference between N2 and N1 was smaller (*M*s = 43.30 and 40.63).

The analysis revealed also a main effect for ability to cope with stress in predicting mental load (p=.01). People that were able to cope with stress more efficiently experienced less mental load (*M*=8.43) than people that are not so good at coping with stress (*M*=13.23). Analysis also showed an interaction between Ability to Cope with Stress and Task Difficulty in predicting mental load. The difference between difficult and easy task was smaller among people that are good in coping with stress (*M*=8.64 and 8.22) than among people that are not as good in coping with stress (*M*s = 14.11 and 12.34).

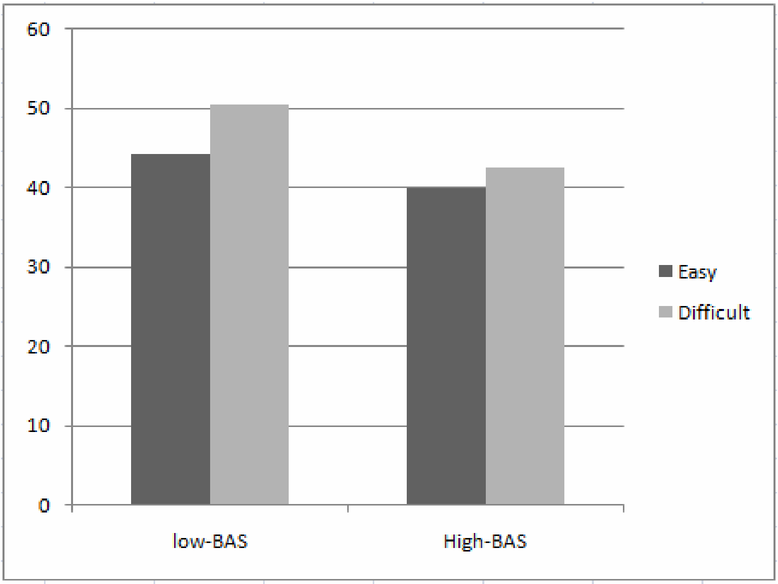


Fig. 2. Workload scores for low and high BAS scorers during easy and difficult task

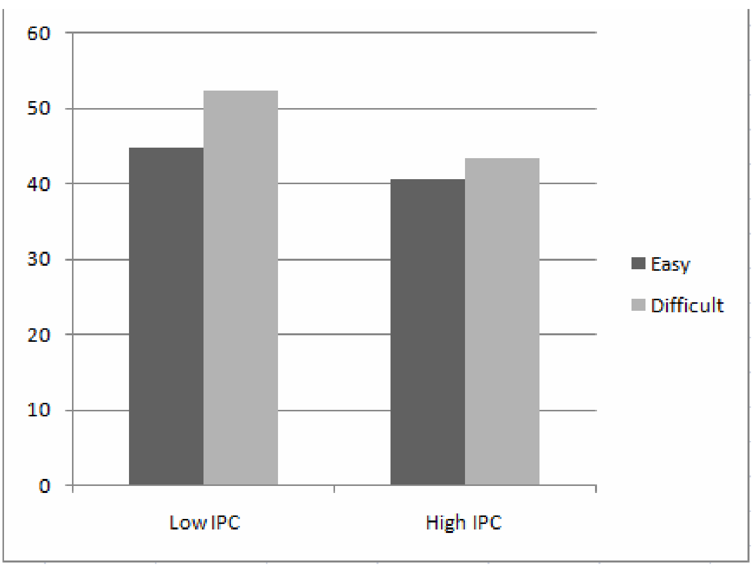


Fig. 3. Workload scores for participants scoring low and high on information processing capacity (IPC) during easy and difficult task

3.1 Reaction Time

In regard to task performance, the LMM analysis revealed a significant interaction between BIS and Task Difficulty in predicting reaction time to tactual stimuli (p<.001). For low BIS scorers the reaction time to the tactual stimuli was considerable higher in difficult than easy task (*M*s = 1.36 and 1.22), whereas for high BIS scorers the difference in reaction time in easy and difficult task was smaller (*M*s = 1.28 and 1.19).

In predicting reaction time, the analysis also revealed a significant main effect for ability to cope with stress (p<.001). Participants scoring low on ability to cope with stress reacted faster to the tactual stimuli than participants scoring high on ability to cope with stress (*M*s = 1.25 and 1.27).

Analysis also showed a significant main effect for information processing capacity (p <.001) and an interaction between Task Difficulty and Information Processing Capacity (p = .001) in predicting reaction time. Reaction time was lower among participants scoring high on information processing capacity than among participants scoring low on information processing capacity (*M*s = 1.26 and 1.27). As illustrated in Figure 4., the difference between reaction times in difficult and easy tasks was greater among subjects scoring low on information processing capacity (*M*s = 1.36 and 1.18) than among subjects scoring high on information processing capacity (*M*s = 1.31 and 1.21).

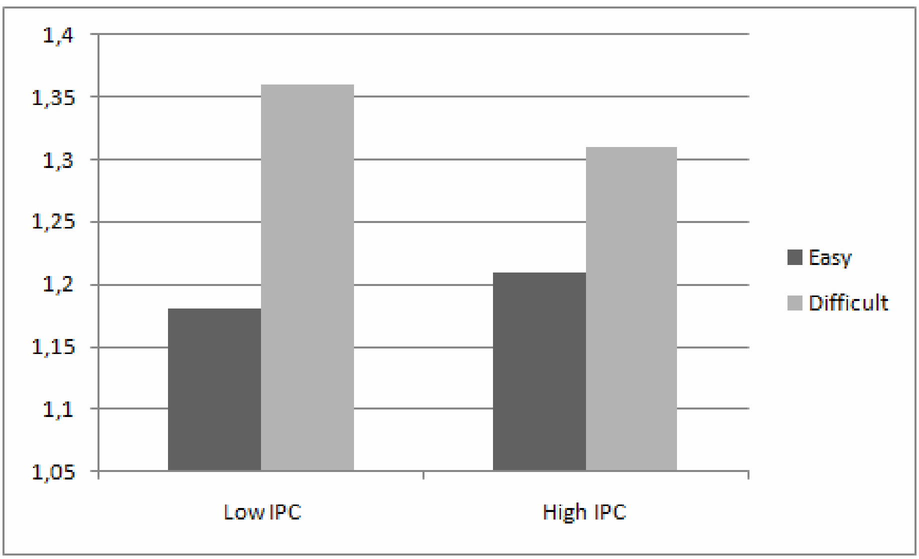


Fig. 4. Reaction times during easy (N1) and difficult (N2) tasks among low and high information processing capacity scorers

4 Discussion

The results showed several individual differences in workload while doing multitasking with a computer. We found among other things that the difficult task was especially more difficult than the easy task for participants scoring low on information processing capacity, whereas participants scoring high on information processing capacity performed almost equally well in easy and difficult task (in terms of self-report work load). Similarly, people with lower capacity to cope with stress tend to feel more overall mental load than people with higher capacity to cope with stress. The result on task performance showed that in addition that people with high information capacity tend to evaluate easy and difficult tasks as almost equally loading, they also performed almost equally well in both conditions: the difference between reaction times in difficult and easy tasks was significantly lower among subjects scoring high on information processing capacity than among subjects scoring low on information processing capacity. In addition, in general reaction time was lower among participants scoring high on information processing capacity than among participants scoring low on information processing capacity. Thus the results suggests that information processing capacity can be measured meaningfully with a simple likert scale, and that the scores on this measures can be used to predict the participant's processing capability (reaction time and work load) in a multitasking task environment.

We also found that among participants scoring high on BAS there was not a large difference in workload ratings between easy and difficult task, whereas the opposite was true for low BAS scorers. We also found that for low BIS scorers the reaction time to the tactual stimuli was considerable higher during difficult than easy task, whereas for high BIS scorers the difference in reaction time during easy and difficult task was smaller. As low BIS correlates with high BAS and low BAS correlates with high BIS, it seemed that even though the results supported the idea that extrovert subjects are prone to any kind of stimuli as opposite to low BAS or high BIS scorers, their performance during easy and difficult task nevertheless differed greatly. However, these speculations and reasons for these findings need to be validated in follow-up studies.

In regard to the reaction time, we also observed that participants scoring low on ability to cope with stress reacted faster to the tactual stimuli than participants scoring high on ability to cope with stress. We feel that this finding reflect the higher reactivity of the people that are not very good in coping with stress on external stimuli on their body (a possible stressful stimuli).

In sum, the results show that there are many individual factors that may affect the nature and magnitude of work load and task performance in multitasking computer environments. The results have implications for the design of multimodal systems as clearly some people can perform quite heavy tasks easily while for others they may be more difficult. Optimally a multitasking environment should be designed so that it would be able to adjust it to adapt the task interface and the presentation of information to fit the user capabilities.

Acknowledgments. This study was supported by the Finnish Funding Agency for Technology and Innovation (TEKES) project "Supporting Situation Awareness in Demanding Operating Environments through Wearable Multimodal Interfaces (SAWUI)".

References

1. Carver, C.S., White, T.: Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. Journal of Personality and Social Psychology 67, 319–333 (1994)
2. Kallinen, K., Ekman, I.: Taktiili-informaation esittäminen matalan ja korkean tason auditiivisen ja visuaalisen kuormituksen aikana: empiirisiä tuloksia. Sawui projektin raportteja (12 sivua) (2010)
3. Laarni, J., Heinilä, J., Häkkinen, J., Kalakoski, V., Kallinen, K., Lukander, K., Löppönen, P., Palomäki, T., Ravaja, N., Savioja, P., Väätänen, A.: Supporting Situation Awareness in Demanding Operating Environments through Wearable User Interfaces. In: Harris, D. (ed.) EPCE 2009. LNCS, vol. 5639, pp. 13–21. Springer, Heidelberg (2009)

Finding a Relationship between Internet Anxiety and Human Behavior

Santosh Kumar Kalwar, Kari Heikkinen, and Jari Porras

Lappeenranta University of Technology,
Department of Information Technology, Lappeenranta, Finland
{Firstname.Lastname}@lut.fi

Abstract. This chapter introduces a relationship between Internet anxiety and human behavior. The Internet has become an avenue for service provision, third-party applications, connectivity and social media. Thus, a clear understanding and comprehensive description of the relationship between Internet anxiety and human behavior becomes essential. The objective of this paper is to provide an assessment of the behavior and anxiety patterns of different age groups on the Internet. The study contributes to knowledge on human behavior and Internet anxiety. The preliminary investigation was realized with a small number of participants using qualitative research methodologies, including a structured themed interview. The result indicates that participants felt more anxiousness on Internet time delay anxiety, Experience anxiety, and Net Search anxiety.

1 Introduction

"*A man who carries a cat by the tail learns something he can learn in no other way*."
--Mark Twain

With reference to above quoted sentence, imagine the world of the Internet as if it were a cat. It is understood that a user's experience on the Internet might be pleasant or unpleasant. Thus, a user who uses contents on the Internet also "*learns something that he can learn in no other way*". Regardless of millions views of 'cat videos' found in YouTube, thousands of friends created/destroyed in the Facebook, and countless contents spawning on the Internet daily, it is still extremely difficult to reduce and measure user's behavior on the Internet. Moreover, one could argue that everybody wants to be on the Internet regardless of race, color, country, ethics and culture. Thus, the Internet can be seen from many different perspectives. On positive note, contents on the Internet acts as a great platform to share, collaborate and learn new stuff. On negative note, consuming lots of unnecessary information could cause anxiousness and stress. It seems that we no longer live in the 'age of information' but we 'float in the age of information'. Therefore, Internet anxiety is a serious problem simply because it causes cultural, social, economical and psychological stress. We might define Internet Anxiety as the state of helplessness, a conscious struggle for right information at the right time, and a feeling of distress to some extent. Similarly, a user's action or reaction at the Internet interface could define his/her behavior.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 359–367, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Finding a relationship between user's behavior and Internet anxiety seems significant in order to 'axe the trouble' of using contents on the Internet.

2 Related Work

In twenty years of research, many researchers on HCI (Human Computer Interaction) field have reported about the computer anxiety [8, 9, 10, 11, and 12]. More elaborately, Compeau, D.R., et al. proposed a model based on Bandura's social cognitive theory in which longitudinal study with 394 participants were performed and reported, "self-efficacy and outcome expectations impact on an individual's affective and behavioral reaction to IT" [8]. Another study with Durndell, A. and Hagg, Z. suggests gender effects on computer usage –"males with higher computer self-efficacy, lower computer anxiety, and positive attitudes towards the Internet use than females" [9]. While Heinssen, R.K.,et al. reported no significant difference in gender for computer anxiety[10].In another study, Mcilroy, D., et al., demonstrated students with "computer phobia" on the basis of self-report measures [11]. Popovich, P.M., et al., has shown that there is no gender differences towards the use of computers, computer courses in college, amount of time spent using computers or degree of self-reported computer anxiety [12].

Twelve years ago, one study suggested that Internet users used to spend most of the time online using email and Web [7]. It has also been hypothesized that those who have higher computer anxiety will also possess higher Internet anxiety [6]. On the other hand, Internet anxiety is related to computer anxiety [1]. In very simple terms, an anxiety caused by contents on the Internet can be defined as the Internet anxiety. Empirical data has illustrated that computer anxiety and Internet anxiety vary [2, 3, and 4]. However, it is understood that Internet can cause anxiety [5]. Many studies have elaborately discussed on Internet Addictive Behavior (IAB). These behaviors encompasses physiological factors such as depression, anxiety, loneliness and personality factors [13], gender and Internet usage [14], self esteem [15] and personal attachments [16]. Nevertheless, none of the existing literatures is focused on assessment of human behavior and anxiety on the Internet.

With an advent of services, applications, games, social networking sites on the Internet, many things have changed in recent years. There are evidences of anxiety and stress caused by online contents. However, Literature study suggests that it is very difficult to determine the self-reinforcing relationship between anxiety and behavior. Moreover, in literature there is no link between anxiety and behavior.

3 Research Questions and Methodology

In this chapter, we mainly focused on building relationship between human behavior and Internet anxiety. Figure 01 below illustrates how the Internet anxiety and human behavior study was conducted. Two sessions of Interviews (Interview 1 and Interview2) were conducted with varying number of users. A set of questionnaires (Q1 and Q2) were prepared and distributed among users.

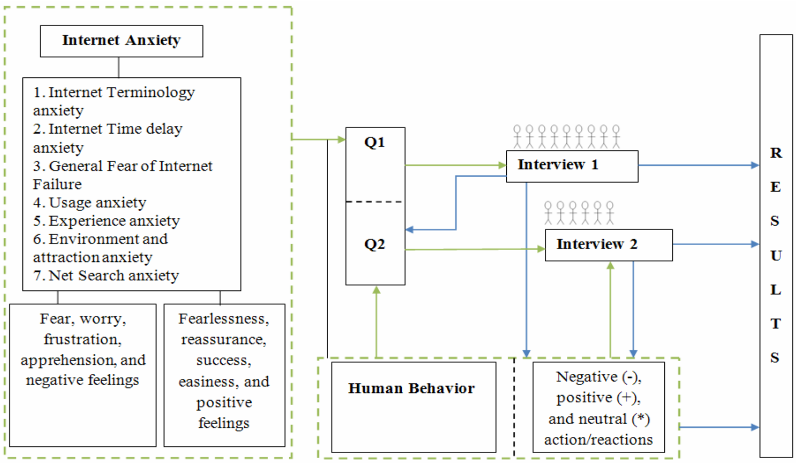


Fig. 1. Diagram illustrating the process of the study, the green color indicates the input/literature whereas the blue color indicates the output/results

What is the nature of Internet anxiety? The below bulleted points summarize the assumptions about the Internet anxiety. Is any part of this wisdom correct? The various literature reviews on Internet anxiety and human behavior was taken into account. The demographic data with professors, researchers and students were collected during this research enhances our understanding of the phenomenon and may help to break down some of the wisdom surrounding myths on Internet anxiety.

- Majority of young students are not anxious about using the contents on the Internet.
- Students with high technical skills can easily understand almost anything on the Internet.
- Most human behave normally on the Internet.
- Women will have higher level of Internet anxiety problems.
- Any information can be obtained from the Internet therefore, there is no anxiety.
- User/human with greater computer experience will suffer less Internet anxiety.
- Social networking causes no anxiety.
- I could never get depressed—I am too strong a person on the Internet.
- My Interaction with the Internet do not cause any harm to others.
- I do not care about the Internet anxiety and its types.
- I am never anxious. My online behavior is excellent.
- All types of Internet anxieties are caused by … (A single cause is referenced. For example. Personality, past experiences, health, social-economical scenarios, mental illness, fear and apprehension.)

Can Internet anxiety be treated? If Internet anxiety is real, than the treatment is necessary. Until date, Internet anxiety has not been realized as "general anxiety". The necessary approaches for the treatment of Internet anxiety are vital. In case of human behavior on the Internet, there are wide ranges of questions that need to be resolved if the problem has to be fully understood:

What is the human behavior on the Internet? One needs to conceptualize behavior in the context of Internet usage, task and user's characteristics. Can human behavior be measured? If human behavior can be measured, the types of measurement should be specified. Various types of research methodologies need to be discussed. After the investigation of these sub-questions, we came across the main question:

How can we build a relationship between human behavior and human anxiety on the Internet? To build any kind of relationship, the correlation of two variables is necessary. The two variables mean 'Internet anxiety' and 'human behavior'. One needs to investigate these two variables and their types. Thus, the coupling between Internet anxiety and human behavior has been examined with small user groups in different demographics conducting interviews. The data was gathered by Qualitative research approach such as Interview analysis. There were two sessions of Interview conducted among participants. In the first session of the Interview, first round of the questionnaires were administered to nine participants with different demographics. All the nine participants accepted the interview questions that yielded fruitful response which was recorded using audio tapes and evaluated

4 Study Setup

The aim of the study was an attempt to put the human in different age, gender and Internet context. These user's opinion data was audio taped. Hence, we started to design the study by conducting a first session of Interview with only nine participants in the laboratory settings. In session one, the interview was conducted with three smaller groups consisting of three participants each. There were two reasons behind conducting this session. Firstly, we wanted to make sure that our study set-up was clearly understood by the participants. Therefore, we needed some time to define types of users and their demographics. Secondly, we believed that it is important to start with smaller number of participants. The participants in the first session were all male. The participants in the second session were male and females. The three participants from each group were divided such that each group age range was different. The first group consisted of professors, second group consisted of researchers and third group consisted of students.

Sample evaluation and analysis. The sample consisted of fifteen participants. These samples were divided into three groups based on age. Each group consisted of three equal participants. It is important to note that three participants were above the age of 40 and they belong to the category of professors or researchers. They all have used the Internet since 1988 or 1989. The mean average age of all the participants has been calculated as 31, 89.

Interview Questions. The interview was structural interview. The reason of conducting this interview was to select the expert participants who had background knowledge about the Internet technology in general. The Interview began with simple questions and later difficult questions were given to the participants.

Interview Questions 01

1-4: Why/when/where/how, you are using the Internet.

5: What Internet means to you?
6: What is not in the Internet that you would like to have?
7: What is in the Internet that you would like to use offline?

Interview Questions 02

1. What are categories of problems that you encounter on the Internet?
2. How are you feeling, when you encounter such problems?
3. What sorts of thing affects your feeling?
4. How are you behaving, when you encounter such problems?
5. What sorts of things affect your behavior?
6. In above seven types of anxieties, which one you feel the most and why? *(Strongly Disagree: 1, Disagree: 2, Neutral: 3 Agree: 4, Strongly Agree: 5)*
7. What are other types of anxieties you have encountered on the Internet?
8. What are the technical-enablers to overcome your anxiety (e.g., plug-ins in browser, advertisements blocker client programs etc)?

Data Collection. The data was collected by audio tape recording and note taking mechanism. The entire interviews were recorded and notes were taken at the same time. After the completion of interviews with all the participants, annotation of audio tape and notes were carried out. The summary of that review is discussed in result and discussions section.

5 Results and Discussions

We divided our results based on participants understanding about the Internet in general by asking "W/H questions". After the interview for the session was finished, the tape-recorded was analyzed. Let us start by describing what happened with participant 2 and 3 on seven different interview questions. P2 is 24 years old student, using the Internet since 1998 and P3 is 30 years old researcher using the Internet since 1997. When they both arrived in the lab, they said that they felt quite relaxed. The study set up was demonstrated and demographic details of paper were given. The detailed transcript of the interview is avoided. Likewise, the participants 9 is a researcher who works in the laboratory. *He has been using the Internet since 1996 and is 31 years old*. This participant usage the Internet "*quite a lot*" and "*getting information*" while at work or home is very significant for him. The common keywords recorded from his interview have been shown in the figure below.

Afterwards, the interview was conducted individually with those participants who were above 40. *The participant 8 is a professor and he has been using the Internet since 1988*. The participant uses the Internet "*whenever, wherever*". For this participant, the Internet is "*possibility*" for various "*information*" and believes "*everything is on the Internet*". Similarly, the interview was conducted individually with the participant 1. *He is an adjunct professor and he has been using the Internet since 1989.* The participant uses the Internet "*a lot*", and mainly at "*work*" or at "*home*". For this participant, the Internet is "*information*", "*connectivity*" for various "*services*" and believes "*synchronization of calendar*" is not present on the Internet. The study was conducted among three participants 4, 5 and 6 who are students studying in the technological university. They have been using the Internet since 2001, 2000 and 1999 respectively. These students view the Internet as the "*information source*" and they want the services on the Internet to be "*free*". They extensively use the *multimedia contents* on the Internet. In addition, they also want to learn, share and collaborate with other users using social networking sites (SNS) for example, Facebook, Twitter and MySpace. They use the Internet with their laptops, desktop and mobile devices. The final interview was carried out with participant 7. *He is a researcher who works in the laboratory. He has been using the Internet since 1989.* He uses the Internet almost "*every time*" and for him the Internet means "*connectivity*" sharing of various "*services*", "*information*", different types of protocols for communications. *He believes that "calendar" synchronization is not present in the offline mode and it would be useful application.*

It can be concluded that all the participants viewed the Internet as the medium for communication, sharing contents, information, and connectivity. These participants were asked about their applicability of Internet and their common reply was "*almost everywhere*". The majority of participants did not use the Internet using the mobile devices. Two participants suggested "*calendar synchronization*" on the Internet would be a great idea. The student participants suggested mainly on "*free services*" and "*free contents*" (for example, Online TV, Free Streaming Multimedia Contents and Free WebPages). They would like to share contents with friends and family "*almost every time*".

In response to the sixth item of the Interview Question 02, following table and graph was recorded from the seven users. As there has been no threshold to define Internet anxiety, we presumed 3.5 to be more anxiousness (*See Table 1 and Figure 02*). Therefore, the tabular data below shows that, Internet Time delay anxiety, Experience anxiety and Net search anxiety are more in the users than others types of anxieties.

Table 1. Rating of seven different types of Internet anxiety by users

Types of Internet anxieties	User1	User2	User3	User4	User5	User6	User7	Average
Internet Terminology anxiety	1	2	3	2	1	3	2	2
Internet Time delay anxiety	5	4	5	2	4	5	5	4,28571
General Fear of Internet Failure	3	4	4	1	3	4	2	3
Usage anxiety	3	3	3	4	4	2	3	3,14286
Experience anxiety	4	5	3	4	3	3	4	3,71429
Environment and attraction anxiety	2	4	2	5	4	4	3	3,42857
Net search anxiety	4	4	4	3	3	3	4	3,57143
Average	3,14286	3,71429	3,42857	3	3,14286	3,42857	3,28571	

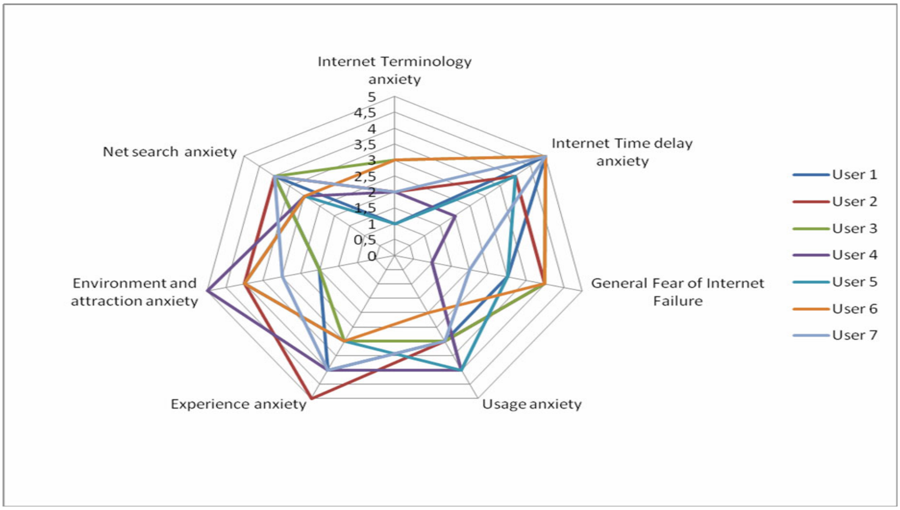


Fig. 2. The graph illustrating seven different types of users and their ratings for severity of anxiousness

5.1 Relationships between Anxiety and Behavior

The following discussion based on the themed interview resulted into mathematical derivation of relationships between Internet anxiety and human behavior. The process below shows how Internet Anxiety (IA) could be linked to human behavior (HB). This is because a user while considering use of the Internet may feel or inhibit emotions related to Internet use that might lead to anxiety. A set of assumptions for the assessing relationship between anxiety and behavior has been developed. In the figure below, Internet Anxiety is abbreviated as 'A' and human behavior is abbreviated as 'B'. The path of the Internet Anxiety and human behavior are depicted into four smaller diagrams. For example, In a) anxiety is shown as a dynamic entity whereas in c) behavior is shown dynamic. Similarly, in b) both anxiety and behavior are linked to each other. In d) both anxiety and behavior are shown with various causes separately (*See Figure 03*).

1. A ≠ B
2. IA ≠ HB or, IA↔ HB
3. IA ~ HB means HB ∈ [IA] (and equivalently IA € [HB])
4. IA € {fear, worry, frustration, apprehension, and negative feelings}
5. HB= {negative action/reaction}
6. IA € {fearlessness, reassurance, success, easiness, and positive feelings}
7. HB= {positive action/reaction}
8. IA ∉ {fearlessness, reassurance, success, easiness, and positive feelings} and
9. ∉ {fear, worry, frustration, apprehension, and negative feelings}
10. HB= {neutral action/reaction}
11. A= (fear, worry, frustration, apprehension, and negative feelings)

12.B= (fearlessness, reassurance, success, easiness, and positive feelings)
13.C= (no feelings/emotions)
14.IA= (A, B, C)
15.HB= (negative, positive and neutral action/reactions)
16.If negative=- , positive= + and neutral = * then {A, B, C} = {-, +, *}
17.IA $\bowtie$ HB

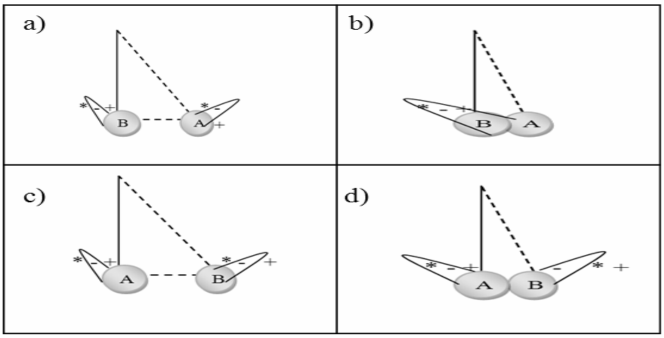


Fig. 3. The diagram shows the relationship between anxiety and behavior

6 Conclusions and Future Work

From our preliminary study, the Interview result suggests that the Internet anxiety could be part of human behaviour but not vice versa. Furthermore, among seven types of anxiety identified, participants felt most anxious on *Internet time delay anxiety*, *Experience anxiety,* and *Net Search anxiety*. Majority of the participants also suggested that they do not have any impact of anxious behaviour on the Internet. A relationship between Internet anxiety and human behaviour can only be determined by understanding user's behaviour. In order to understand how human/user behaves on the Internet it is extremely important to measure user's reaction at the Internet interface. Moreover, a user with certain types of Internet anxieties is much easy for measuring behaviour. A priori to understand and find relationship between Internet anxiety and human behaviour is determinant of not only types of Internet anxieties but also user's action/reaction at the Internet interface.

The limitation of study suggests that analyzing and conceptualizing human behaviour may not be the complete solution. Thus, the possible alternative to reduce human anxiety on the Internet is by classification of anxieties and by use of qualitative methodologies to enhance sound empirical results. The future work would comprise of measuring user's reaction to reduce Internet anxieties, building conceptual framework to assess human anxiety on the Internet and creating tools/techniques/methods to measure human anxiety on the Internet. In short, development of technological tools to understand relationship between Internet anxiety and human behavior would be an asset to the community of HCI.

References

1. Thatcher, J.B., Perrewe, P.L.: An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. MIS Quarterly 26(4), 381–396 (2002)
2. Brown, S.A., Fuller, R.M., Vician, C.: Who's afraid of the virtual world: anxiety and computer-mediated communication. Journal of the AIS 5(2) (2004)
3. Thatcher, J.B., Loughry, M.L., Lim, J., Harrison McKnight, D.: Internet anxiety: An empirical study of the effects of personality, beliefs, and social support. Information & Management 44(4), 353–363 (2007) ISSN 0378-7206, doi:10.1016/j.im.2006.11.007
4. Brown, S.A., Fuller, R.M., Vician, C.: Who's afraid of the virtual world: anxiety and computer-mediated communication. Journal of the AIS 5(2) (2004)
5. Macaulay, M.: The effects of web-assisted learning on students' anxiety. Journal of Educational Computing Research 28(3), 221–230 (2003)
6. Davis, R.A.: A cognitive-behavioral model of pathological Internet use. Computers in Human Behavior 17, 187–195 (2001)
7. Young, K.S.: Internet addiction: The emergence of a new clinical disorder. CyberPsychology & Behavior 1, 237–244 (1998)
8. Compeau, D.R., Higgins, C.A., Huff, S.: Social Cognitive Theory and individual reactions to computing technology: A longitudinal study. MIS Quarterly 23(2), 145–158 (1999)
9. Durndell, A., Hagg, Z.: Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an Eastern European sample. Computers in Human Behavior 18, 521–535 (2002)
10. Heinssen, R.K., Glass, C.R., Knight, L.A.: Assessing Computer Anxiety: Development and validation of the Computer Anxiety Rating Scale. Computers in Human Behavior 3, 49–59 (1987)
11. Mcilroy, D., Sadler, C., Boojawon, N.: Computer phobia and computer self efficacy: their association with undergraduates' use of university computer facilities. Computers in Human Behavior 23, 1285–1299 (2007)
12. Popovich, P.M., Gullekson, N., Morris, S., Morse, B.: Comparing attitudes towards computer usage by undergraduates from 1986 to 2005. Computers in Human Behavior 24, 986–992 (2008)
13. Yu, S.C.: Internet using behaviors, Internet addiction, and related factors among university students. Master thesis, Taiwan (2001)
14. Lin, S.S.J., Tsai, C.C.: Sesation seeking and Internet dependence of Taiwanese high school adolescents. Computers in Human Behavior 18, 411–426 (2002)
15. Hsau, M.: Research on personal characteristics, the behavior of using Internet and Internet addiction for Taiwanese college study. Master thesis, Taiwan (2000)
16. Wang, Cheng Wha: The personality, interpersonal interaction on the cyberspace and the Internet addictive behavior. Master thesis, Taiwan (2000)

Human Behavioral Simulation Using Affordance-Based Agent Model

Namhun Kim[1], Jaekoo Joo[2], Ling Rothrock[3], Richard Wysk[4], and Young-Jun Son[5]

[1] Design and Human Engineering, Ulsan National Institute of Science and Technology, Republic of Korea
[2] Systems and Management Engineering, Inje University, Republic of Korea
[3] Industrial and Manufacturing Engineering, the Pensylvania State University, University Park, PA 16802, USA
[4] Industrial and Systems Engineering, North Carolina State Universty, Raleigh, NC 27695, USA, and
[5] Systems and Industrial Engineering Department, The University of Arizona, Tucson, AZ 85721, USA
nhkim@unist.ac.kr

Abstract. In this paper, we propose a novel agent-based simulation modeling of human behaviors. A conceptual framework of human behavioral simulation is suggested using the ecological definition of affordances in order to mimic perception-based human actions interacting with the environment. A simulation example of a 'warehouse fire evacuation' is illustrated to demonstrate the applicability of the proposed framework. The perception-based human behaviors and planning algorithms are adapted and embedded within human agent models using the Static and Dynamic Floor Field Indicators, which represent the evacuee's prior knowledge of the floor layout and perceivable information of dynamic environmental changes, respectively. The proposed framework is expected to capture the natural manners in which humans participate in systems and enhance the simulation fidelity by incorporating cognitive intent into human behavior simulations.

Keywords: Human Behavior, Affordance Theory, Finite State Automata, Agent-based Modeling, and Simulation.

1 Introduction

While systems theory has grown rapidly, modeling and simulation of human-involved systems has not kept pace. We suspect that it is, in part, because no common framework exists to accommodate both human cognitive models and the discrete system representations within which humans play critical roles. This creates a major modeling void as most complex systems contain human activities. In this research, we present a generic simulation framework to fill up the void using the ecological concept of affordance and a formal modeling methodology for human-involved complex systems.

In 1979, Gibson initiated an ecological understanding of human actions in animal-environment systems [1]. According to him, a human action is regarded as a consequence

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 368–377, 2011.
© Springer-Verlag Berlin Heidelberg 2011

of the juxtaposition process of affordance and effectivity (an individual's property of ability to take a specific action) which are directly perceived by the human. That means a human makes decisions to take actions based on perceived information regarding sets of affordance-effectivity couples.

From this perspective, we run into the challenge to address the human in the system and to provide a modeling vision for a human-involved system. In this research, an affordance-based simulation framework that is capable of analyzing perception-based human behavioral patterns in complex and uncertain environs is developed for modeling systems with people. For illustrative purposes, we discuss the problem from the standpoint of the ecological concept of affordance to interpret human actions and simulate their perception-based planning and behavioral patterns. The proposed framework will provide a predictive analysis capability for the design of human-involved system.

The remainder of this paper is organized as follows: Section 2 provides the related work and background of this research. Section 3 presents the conceptual simulation framework and its components to cater for human-involved discrete event simulations. We demonstrate the proposed simulation framework by implementing the 'ware house fire evacuation' scenario in which human evacuees make decision based on their perception as well as previous knowledge in Section 4. Finally, we conclude this paper with a discussion of the direction for future research in Section 5.

2 Background

2.1 Human Behavior Modeling and Simulation

Human decision behaviors have been studied by researchers in various disciplines such as artificial intelligence, psychology, cognitive science, and decision science. Lee et al. [2] classified human decision behavior models into three major categories based upon their theoretical approach: 1) economics-based approach, 2) a psychology-based approach, and 3) a synthetic engineering-based approach. Each approach exhibits strengths and limitations. First, models employing the economics-based approach have a concrete foundation, based largely on the assumption that decision makers are rational [3, 4]. However, one limitation is the inability to represent the nature of human cognition. To overcome this limitation, models using a psychology-based approach have been proposed [5, 6]. While these models explicitly account for human cognition, they generally address human behaviors only under simplified and controlled laboratory conditions [7].

While current synthetic models available in the literature consider human perceptions explicitly, none of them is grounded both on the ecological concept of affordance and a formal system theoretic method for human-involved complex systems. Therefore, the goal of this research is to develop a generic framework for affordance-based simulation of human behavior using affordance-based finite state automata (FSA).

2.2 Affordance Theory and Perception-Based Human Actions

Gibson first defined the affordances of the environment as what it offers to the animal (person), what it provides or furnishes, either for good or ill [1]. Since Gibson proposed his definition, the notion of affordance has been further refined and theorized. For example, Turvey presents a perspective on the ecological ontology of affordances with links to prospective control [8]. Turvey bases the definition of affordance in terms of properties that represent a potential state and are not currently realized (called dispositional properties or dispositions). Dispositions occur in pairs in which a property of the environment (i.e., walk on – ability for the person) is complemented by a property of the animal's capability known as an effectivity (i.e., to walk on the stairs' surface). So the terms of affordance and effectivity can be combined together so that they incur a different property (i.e., climb the stairs) to be activated [8]. Specifically, Turvey presents a formal definition of affordances mathematically using a juxtaposition function as follows;

Let $W_{pq}=j(X_p,Z_q)$ be a function that is composed of an animal (X) and a environmental object (Z), and further p and q be properties of X and Z, respectively. Then, p refers to an affordance of X and q is the effectivity of Z, if and only if there exists a third property r such that:

i) $W_{pq}=j(X_p,Z_q)$ possesses r,
ii) $W_{pq}=j(X_p,Z_q)$ possesses neither p nor q, and
iii) Neither X nor Z possesses r, where r is a joining or juxtaposition function.

For example, in case of a person-climbing-stairs system (W), a person (Z) can walk (q), stairs (X) can support something (p), and they together yield climbing property (r). This formal definition of affordance, effectivity, and juxtaposition function can be mapped to the precondition set of state transition function in the FSA and provides a foundation that the concept of an affordance can be combined with software engineering and systems theory.

2.3 Affordance-Based Formal Modeling of Perception-Based Human Actions

The affordance-based modeling approach has been widely adapted for designing robot controls and mimicking human actions in specific environmental situations [9]. However, most of these efforts lack formal ways to model human actions with respect to system transitions, due to the low level of abstraction in modeling perceptual properties.

Kim et al. suggest an affordance-based descriptive formalism for human-involved complex systems using finite state automata. In their work, an environmental system is defined as a set of nodes and arcs which describe discrete states of the system and the transitions between states, respectively [10]. Then, a set of transitions between states is triggered by certain human actions to lead to the next states. Affordance-effectivity combinations are considered as preconditions for actualizing possible human actions. In their work, a perception-based human action and its associated properties of an affordance and effectivity are systematically defined as placeholders in the context of discrete systems for implementation of a dynamic juxtaposition process (see Figure 1). There are two kinds of state transitions: external state transition and internal state transition. The external state transition can bring physical changes of system state caused by a specific human action in the human-environment system. On the other hand, the internal state transition is concerned with a precondition(s) that must be satisfied for actuating the external state transition in the system. It connects two sub-states each of which is a combination of a specific affordance and effectivity. These affordances and effectivities are time-varying properties which are determined by the physical situations of environmental elements and human capabilities. In a certain amount of time, t_{int}, if the current status of affordances and effectivities (p_i, q_j) may be changed to (p_m, q_n) to meet a specific action condition, then the juxtaposition function will generate a human action and then an external state transition will occur, and then a physical state of the system goes to the next. The formal modeling of human behaviors in the sense of perception-based decision making provides a systematic way to include physical preconditions of human action possibilities within computable modeling methodologies.

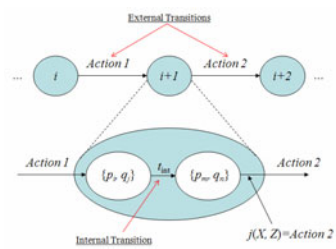


Fig. 1. External and internal transitions of Affordance-FSA formalism [10]

3 Conceptual Framework of Affordance-Based Human Behavior Simulation

3.1 System Architecture of Affordance-Based Simulation

In the system theoretic perspective, a human can be regarded as a non-deterministic and unpredictable agent within an environment. Envision of theory of affordance allows us to include the properties of perception-based actions into discrete event-based systems and to build the affordance-based simulation framework [10]. The proposed affordance-based simulation framework is able to evaluate action possibilities and predict human behaviors using the embedded formal automata models of human-involved systems. It considers both affordance and effectivity as dynamic control parameters which trigger human actions in the simulation in order to deal with the interactions between human and dynamic environmental elements.

Figure 2 depicts the system architecture of the proposed affordance-based simulator, where four major components include 1) Affordance-based FSA model, 2) Agent Models, 3) Event Generator, and 4) Human Action Planner. The Affordance-based FSA models are formal automata models of human-involved systems that describe the whole state map of human actions and dynamics of environments. The FSA model provides dynamic (temporal and spatial) situations and preconditions of possible transitions for agents in the system. The agent models are classified into a human agent and an environmental agent. A human agent represents each human in the system. An environmental agent describes each environmental element (or cue) of the system which may affect human decision makings in the simulation model such as fire, smoke, and flood. Each human agent has an action planning (or decision) algorithm and the environmental agent follows its own governing rule based on physics, while they make transitions from one state to another on the state map of the affordance-based FSA. The event generator triggers each event according to pre-defined human action plans by perceived property data of affordances and effectivities from a human agent. The human action planner generates a feasible sequence of actions reaching from the current toward the goal state for a human agent. In order to do that, the planner refers to dynamic data of affordances and effectivities as well as static data of human prior knowledge such as the layout of an environmental space. If an unexpected or undesired situation occurs in the system, it causes a transition that leads to deviation from the active plan. If it happens, the planner immediately recalculates the plan in order to cope with the dynamic change of the environment.

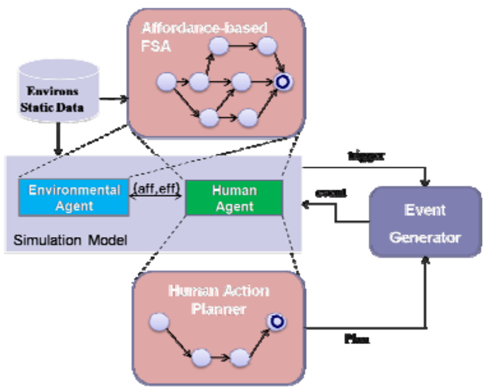


Fig. 2. System architecture of the proposed affordance-based human behavior simulator

3.2 Agent Models

The concept of agent can vary by its application area. In software engineering, the most widely accepted definition of agent model is that: the agent is a computer system that is situated in some environment and capable of autonomous action to meet the system objectives [11]. An agent in simulation model implies the nature of each entity and expresses the complex interaction with other agent in the environment, so that the simple agent rule can generate the complex system behaviors.

In this research, two different types of agent models are defined for the model verification (see Figure 3). While a human agent model is represented by goals, perception abilities, decision making algorithm and action capabilities, an environmental agent maps dynamics of environmental elements onto the system model. The concepts of perception-based human action and planning are used to develop the decision making algorithm in a human agent model in Section 3.3.

Several attributes (e.g. type, name, moving speed, etc) and characteristics for each agent are defined to reflect the diversification of humans and environmental elements in the system. A conceptual model is developed to illustrate functional procedures of and interactions between the human agent and the environmental agent in Figure 4. It is noteworthy that human actions are affected by the information on affordances and effectivities perceived in the human-environment system.

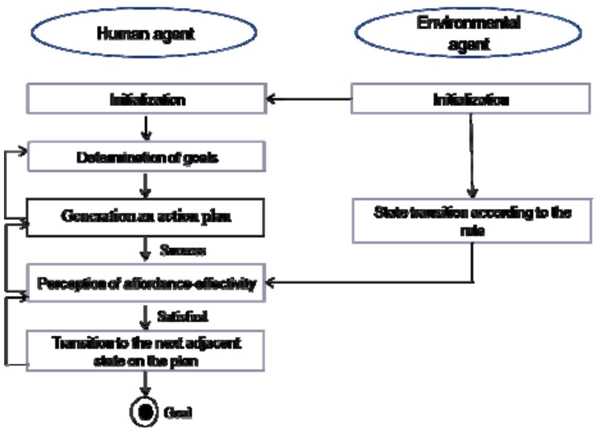


Fig. 3. Conceptual model of human agent and environmental agent

3.3 Human Action Planner

Considering the perception based actions and planning of human agents, we need to define a concept of a perceivable boundary (PB) as a horizontal space that a human can visually perceive surrounding situations. Since perceivable boundary is determined by the position of a person and dynamics of other environmental conditions such as lighting and fog, it should be continuously updated according to the movements of a person. While the shape and range of a perceivable boundary are dependent on the human visual direction, it is assumed in this example that human agents have 360 degree viewing angle by rotating their body and head of their own free wills.

The plan is generated based on not only static information of human knowledge on the surface layout of an environment, but also perception of dynamic information within perceivable boundary obtained at the very moment of the planning (or re-planning) and decision making. In this research, the static information is defined as Static Floor Field Indicator (SFFI) which is a set of numbers indicating distance to the exit [12], whereas the dynamic information as Dynamic Floor Field Indicator (DFFI) which is a set of numbers representing affordances (move-abilities for a human) of perceived areas within a perceivable boundary.

For the static information in the simulation, it is assumed that a human has previous knowledge about the floor layout regardless of his or her perception of the environment, so as to make a complete plan which persists until achieving his goal. He or she is supposed to know how far the goal position (i.e., the exit) is from the current position in the layout. For the dynamic and complex situation which can change the floor layout, however, it is not the case. Suppose that the evacuation route for human should be changed because of a fire on a corridor within the evacuee's original evacuation route. If the system is too large for a person to perceive the entire area, he or she cannot know every details of a current situation of the environment beyond his or her PB. So, the evacuee with insufficient information of the environment can make his or her evacuation plan only within the perceivable boundary. Beyond the boundary, he or she may be able to make a rough plan with the layout information in his or her memory. The rough plan here means a plan made only with prior knowledge of the layout (static information). On the other hand, the detailed plan is made based on both the static information (prior knowledge) as well as dynamic information (perceptual properties) within the perceivable boundary. The planner generates and updates both types of plans every time it's needed (e.g, when an unexpected situation which blocks the current planning path occurs). While the detailed plan within a perceivable boundary is expressed via affordance-based FSA considering perceptual properties of affordances and effectivities, the rough plan outside of the perceivable boundary is generated based on the FSA model without consideration of the perceptual properties.

The evacuee decides his or her movement to next position based-on his or her objective of behavior in the system which is composed of conditions and goals. An exemplary objective of human behavior is "safe and quick (conditions) escape to an exit (goal)" in the case of emergency evacuations. In this paper, we propose a generic planning algorithm that can be easily applied to generate a human action plan for the affordance-based simulation problem as follow:

- Step1: Define current location of a human agent and perceivable boundary (PB).
- Step2: Based on the current information, make an action plan until the final states (goal state) by calling an objective analysis algorithm. If multiple solutions exist, then randomly choose one of them. (It should be noted that the objective analysis algorithm can be developed for each application domain by specifying its characteristics.)
- Step3: Move the human agent to the next position on the plan as long as the position is currently affordable. If the human agent reaches the final state (goal state), then stop.
- Step4: Update PB and FFIs of cells that newly added to PB. If any state in the plan is not affordable, go to Step 2 and revise the plan. Otherwise, go to Step3.

4 Illustration: Warehouse Fire Evacuation Problem

4.1 Scenario

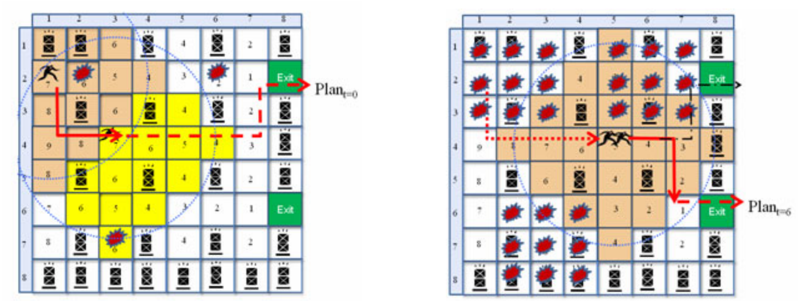
To verify the applicability of the proposed framework of affordance-based simulation, the Warehouse Fire Evacuation (WFE) problem is investigated in this section. In this considered WFE Problem (see Figure 4-(a)), a fire breaks out in a warehouse in which two human operators are working. The warehouse area is equally divided in a rectangular grid of 0.8×0.8 m^2 which is used for either storage or passageway. In this storage area, goods are stacked up so high that operators cannot observe what happens over the storing lots. Fires are firstly broken out at three different locations in the warehouse and are fast propagated to neighboring lots in a certain amount of speed as shown in Figure 4. As soon as an operator perceives the fire, he/she shall have to find a possible and safe route to an exit along passageway by considering perceived surrounding situations in order to escape from the fires. When he/she tries to move to a next passageway lot, if the lot is already occupied with a fire, he/she cannot access to the lot and have to find another passageway that offers an affordance to move. The following assumptions are made to simplify the considered problem:

1. Each evacuee can perceive the situations with 360 degree viewing angle in the warehouse.
2. An evacuee can move to next lots only in the direction of front and back or right and left in right angle. He/she cannot move to diagonal directions in the warehouse.
3. Every evacuee acts independently with each other, but some interactions are possible among evacuees to share their information and knowledge about the floor layout or current situations of the warehouse when they meet in the same places.
4. Fire is only considered as dynamic environmental element. (Smoke and heat are excluded.)

4.2 Model Description

The human behaviors in this evacuation problem can be interpreted as a typical example of affordance-based human actions in prospective control. Turvey defines that prospective control is concerned with future events, usually interpretable as goals to be realized [8]. In this example, each evacuee determines his or her own goal and plan before starting to move. The goal of each evacuee is to get out of the warehouse via one of the exits while avoiding a spreading fire. Once a goal is determined, each evacuee will make a plan for his or her movements along passageways to the exit. To this end, the plan is made by using not only his or her prior knowledge about the floor layout in the warehouse, but perceived information on surrounding situations within the PB at the moment of the planning. A concept of the goal and plan in this evacuation problem is also found in Figure 4-(a).

In order to realize this evacuation problem into a simulation model, the warehouse is described by a two-dimensional grid of equally divided 64 lots. Each cell can be empty, occupied by goods or by fires. Its size corresponds to $0.8 \times 0.8 m^2$, the typical surface occupied by more than two persons in a dense situation [13]. Once the geometry of the floor field and exit locations is determined, each cell is assigned with a constant value of a static floor field indicator (SFFI) representing its static distance to the exits. Nearer the exit, lower values correspond to cells. Thus, these static values indicate to the evacuees the direction to the exit, since they always prefer to move to an adjacent cell with a lower value of SFFI than the current one. So the values of the floor field indicator are used as a basis for generating plans to the way to the exits with dynamic perceptual information on the floor. Cells occupied by goods or fires are assigned with very high values of dynamic floor field indicators (DFFIs) to ensure that there is no affordance of move-ability to an evacuee and he or she will never attempt to occupy them. The floor layout and the values of the static floor field indicators are shown in Figure 4-(a).



(a) human action plan and PB at time t_0 (b) human action plan and PB at time t_6

Fig. 4. Entity state transition model for Warehouse Fire Evacuation Problem

In the affordance-based simulation model, each evacuee is represented as a human agent, and fire is represented as an environmental agent. The continuous time is discredited into time steps which are unit times for the simulation. The mean velocity of evacuees is around 2 m/s [14]. If we set the time step Δt to 0.4 s, human agents move 0.8 m per time step in average. The mean velocity of fire propagation is assumed to be 0.4m/s. This means that a fire broken out in a cell can be propagated into its adjacent cells at every five time steps averagely. Therefore dynamics of the warehouse can be implemented by the floor layout of the warehouse and a dynamic interaction between human agents and propagating fire agents in the warehouse.

To perform this behavior, each human agent perceives whether its adjacent lots provide it with the affordance, "is-move-able" for the human agent, or not. So, in this warehouse evacuation problem, the affordance is "move-ability for an evacuee to an adjacent lot," and the accompanying effectivity for a human agent is "capability to move to an adjacent lot."

5 Conclusion

In this research, we suggest a generic framework for human behavior simulation by combining both human cognitive models and the discrete system representations, which is the affordance-based FSA. The simulation model presented in this paper is agent-based as well as perception-based. The human agent (evacuee) in the demonstrated simulation model has its own algorithm to generate evacuation plans based on both the prior knowledge on the floor layout and what it perceives in the perceivable boundary. The agent model can regenerate evacuation plans whenever its perceivable surrounding environments are changed. This simulates realistic human behaviors which are dependent on perception-based decision making claimed by ecological psychologists.

While we focused on the conceptual framework and human's perception (a major function) in this work, we will consider other human attributes (e.g. social psychology, emotions, cultures, and knowledge levels) in the future. Also, we shall have to validate the simulation results via human experiments in a suitable task environment (e.g. virtual reality).

Acknowledgment. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(20090077993) "

References

1. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
2. Lee, S., Son, Y., Jin, J.: Decision field theory extensions for behavior modeling in dynamic environment using Bayesian belief network. Information Sciences 178(10), 2297–2314 (2008)
3. Mosteller, F., Nogee, P.: An experimental measurement of utility. J. Pol. Econ. 59, 371–404 (1951)
4. Opaluch, J.J., Segerson, K.: Rational roots of irrational behavior: New theories of economic decision-making. Northeastern Journal of Agricultural and Resource Economics 18(2), 81–95 (1989)
5. Edwards, W.: The theory of decision making. Psychological Bulletin 51(4), 380–417 (1954)
6. Payne, J.W.: Contingent decision behavior. Psychological Bulletin 92, 382–402 (1982)
7. Rothrock, L., Yin, J.: Integrating compensatory and noncompensatory decision making strategies in dynamic task environments. In: Kugler, T., et al. (eds.) Decision Modeling and Behavior in Complex and Uncertain Environments, pp. 123–138. Springer, New York (2008)
8. Turvey, M.T.: Affordances and Prospective Control: An Outline of the Ontology. Ecological Psychology 4(3), 173–187 (1992)
9. Kirlik, A., Miller, R.A., Jagacinski, R.J.: Supervisory Control in a Dynamic Uncertain Environment: A Process Model of Skilled Human-Environment Interaction. IEEE Transactions on Systems, Man, and Cybernetics 23(4), 929–952 (1993)
10. Kim, N., Thiruvengada, H., Rothrock, L., Wysk, R.A., Shin, D.: Modeling of Affordances in Human-involved Complex Systems Using Finite State Automata (FSA). IEEE Transactions on Systems, Man, and Cybernetics Part A: Humans and Systems (submitted for publication)
11. Varas, A., Cornejo, M.D., Mainemer, D., Toledo, B., Rogan, J., Munoz, V., Valdivia, J.A.: Cellular automaton model for evacuation process with obstacles. Physica A: Statistical Mechanics and its Applications 382, 631–642 (2007)
12. Zhang, Z., Zhang, C.: Agent-Based Hybrid Intelligent Systems: An Agent-Based Framework for Complex Problem Solving. Springer, New York (2004)
13. Burstedde, C., Klauck, K., Schadschneider, A., Zittartz, J.: Simulation of Pedestrian Dynamics using a two-dimensional cellular automaton. Physica A: Statistical Mechanics and its Applications 295, 507–525 (2001)
14. Daoliang, Z., Lizhong, Y., Jian, L.: Exit dynamics of occupant evacuation in an emergency. A: Statistical Mechanics and its Applications 363, 501–511 (2005)

Investigating the Effects of Metacognition in Dynamic Control Tasks

Jung Hyup Kim[1], Ling Rothrock[1], Anand Tharanathan[2], and Hari Thiruvengada[2]

[1] Industrial and Manufacturing Engineering, Pennsylvania State University, USA
[2] Honeywell Labs, USA
jzk170@psu.edu, lxr28@engr.psu.edu,
{Anand.Tharanathan,Hari.Thiruvengada}@Honeywell.com

Abstract. Metacogntion is a broadly studied concept in cognitive science, educational psychology and developmental psychology. Prior research on metacognition shows that successful learning is often based on specific metacognitive activities which have to be frequently monitored during learning [1]. An important limitation in prior studies is that most experimental tasks were designed within static environments. Only rarely have researchers investigated metacognition during dynamic decision making. Therefore, the purpose of this research is to investigate the significance of metacognition while training to execute dynamic control tasks. Findings from this study demonstrate that the experimental group which was provided more focused feedback had higher situation awareness accuracy, relative to the groups which received relatively less focused feedback or no feedback. In addition, results indicated that metacognitive behaviors which were measured using subjective ratings of Pre and Post Meta probes showed different patterns. Preliminary findings provides evidence that enhanced metacognition has a significant impact on actual situation awareness in dynamic control tasks. Detecting causes for these changes is the next step of the research.

Keywords: Metacognition, Human Performance, Human-in-the-loop-simulation, Situation Awareness.

1 Introduction

Metacognition is thoughts about thoughts or cognition about cognitions [2]. Since Flavell (1976) studied the processes of metacognition, it has been a popular area of study. However, most researchers who have studied metacognition focus on theoretical aspects of metacognition or educational application such as children's learning [3] and academic performance [4]. Transfer of an effective and appropriate strategy from one related problem to another has not received much attention from the metacognitive research community. This transfer is a vital sign which would indicate that a problem solver not only knows how to use a strategy but also when to use it [5]. In this paper, the effects of metacognition and its relation to situation awareness in dynamic control tasks are described. A human-in-the-loop test bed was developed to facilitate the experiment. Rules were embedded into the simulation so that participants can gradually learn

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 378–387, 2011.
© Springer-Verlag Berlin Heidelberg 2011

task-specific rules from training exercises. The primary research question is: *does feedback influence the metacognitive monitoring process and the accuracy of situation awareness in dynamic control tasks?* This study also illustrates how and when a focus on metacognition can be adopted for training complex rule-based model.

2 Literature Review

2.1 Metacognition

There are three facets of metacognition that have been studied widely in the psychology and education fields: metacognitive knowledge, metacognitive monitoring, and metacognitive control. First, metacognitive knowledge is the awareness of one's own cognition. In other words, it is peoples' declarative knowledge about cognitive processes. According to Squire (1986), facts, beliefs, and episodes are three main components of declarative knowledge [6]. Metacognitive knowledge can be generated by the reciprocal action between person and task characteristic with the suitable strategies in a learning situation [2]. Secondly, metacognitive monitoring is the ability to make accurate judgments at the metacognitive level form. Metacognitive monitoring is a necessary component of strategy formation because accurate metacognitive monitoring will produce more effective regulation, and this in turn will produce improved learning [7]. Last facet is metacognitive control which can regulate ongoing cognitive activities such as decision making procedure for using new tactics to solve a difficult problem. In a study conducted by Son (2004), it was shown that participants implemented a metacognitive control strategy to determine the most beneficial re-study option when they had a choice of massing or spacing each item [8]. This is an example of how metacognitive control is utilized to make decision.

Framework for Metacognition. The general framework between metacognition and cognition is conceptualized by the interplay of two levels: meta-level and object-level [9].

In terms of this framework, the definition of object-level is the process of ongoing cognitive activities such as learning and problem solving. The meta-level has a model of subject's understanding of a certain task which is related with ongoing cognitive processes at the object-level. In Figure 1, metacognitive monitoring and control are the main process activities to communicate between two levels. Through metacognitive control process, it is possible to observe the modification at object-level. Information from the meta-level influences the ongoing activities at the object-level. According to Nelson and Narens (1990), monitoring activities must be performed in order to update the model in meta-level [9].

Metacognition in Training Skills. Although there are many extraneous factors that influence metacognitive activities, several studies have shown the effect of metacognitive activities on skills training. According to Teong (2003), outcome from the experimental and case study showed that experimental students who trained on the ability to determine when to make metacognitive decisions outperformed control students on ability to solve word-problems [10]. In addition, results from Veenman's

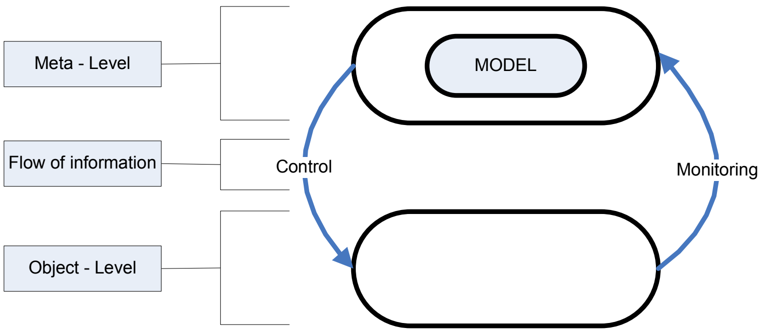


Fig. 1. A framework relation metacognition and cognition level [9]

research (2002) revealed that metacognitive skillfulness was positively related to learning behavior and scores on the qualitative test during inductive training with a complex computer simulation [11]. However, there are still lots of roadblock in metacognition research as the successful transfer of an effective and appropriate strategy from one related problem to another depends on how and when the strategy is used. Moreover, some metacognitive activities showed different contribution to problem solving performance and mental representations across different types of domain areas [12].

2.2 Situation Awareness (SA)

Endsley (1995) defined Situation Awareness (SA) as the perception of the elements in the environment within a volume of time and space (level 1 SA), the comprehension of their meaning (level 2 SA), and the projection of their status in the near future (level 3 SA) [13].

As represented in Figure 2, Level 1 SA is related to perceiving the status, attributes and dynamics of the relevant elements. By using the information, Level 2 SA can be

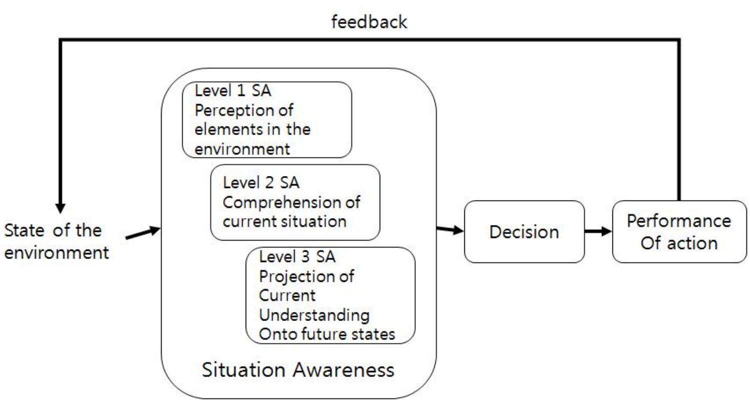


Fig. 2. Model of Situation Awareness [13]

achieved. Level 2 depends on understanding of the significant elements information from Level 1 SA. Achieving Level 3 SA necessitates the use of information from both Level 1 and Level 2 SA. Endsley (1993) applied the model of situation awareness to air-to-air fighter aircraft [14]. Examples of elements include:

- Level 1: location, altitude, and heading of ownship and other aircraft
- Level 2: tactical status of threat aircraft (offensive/defensive/neutral)
- Level 3: projected aircraft tactics and maneuvers

3 Method

3.1 Participants

The total experimentation time took approximately 4 hours per applicant. A total of 48 participants (age 18 or older) were recruited from Pennsylvania State University. All participants were undergraduate and graduate students. Participants were screened for prior experience with the task domain and those who had a past experience of radar monitoring task were excluded from the experiment. This experiment did not exclude any particular subject based on gender, ethnic or racial group. Subject ID was assigned to every subject to preserve their privacy. Participants were randomly assigned to three experimental groups (Skilled-Based Feedback (SBF), Overall Feedback (OF), and No feedback group (CG)). Each group had 16 participants. Average rating of participants' previous experience level using computer interfaces was 3.7 on a scale of 1 (Novice) to 5 (Expert). Average rating of participants' previous experience level playing resource management video game was 2.72 on a scale of 1 (Novice) to 5 (Expert).

3.2 Dynamic Decision-Making Task

The AAWC (Anti Air warfare Coordinator) test Bed is interactive simulation in which a controller must defend his/her ship against hostile aircraft. It was developed at the Human Performance Assessment and Modeling Laboratory at Pennsylvania State University and has been modified for metacognition studies in association with Honeywell. Participants were required to identify unknown tracks and take appropriate action on those aircrafts based on the Rules of engagement (RoE). Participants must focus on identifying unknown aircraft correctly to defend the ship. The details of the Rules of Engagement are shown below:

- Identification Rules (Unknown aircraft only)
  - Make a primary identification of air contact (i.e., friendly, hostile, assumed hostile/friendly)
  - Make an AIR identification of air contact (i.e., Strike, Missile, Helicopter, etc.)
- Warning Rules (Hostile or Assumed hostile only)
  - Issue Level 1 Warning at 50NM
  - Issue Level 2 Warning at 40NM
  - Issue Level 3 Warning at 30NM

- Assign / Illuminate aircraft (at 30NM; Hostile or Assumed hostile only)

AAWC can be run directly from a Windows-based PC. Figure 3 shows a snapshot of the main user interface; the yellow arrows show various components of the interface.

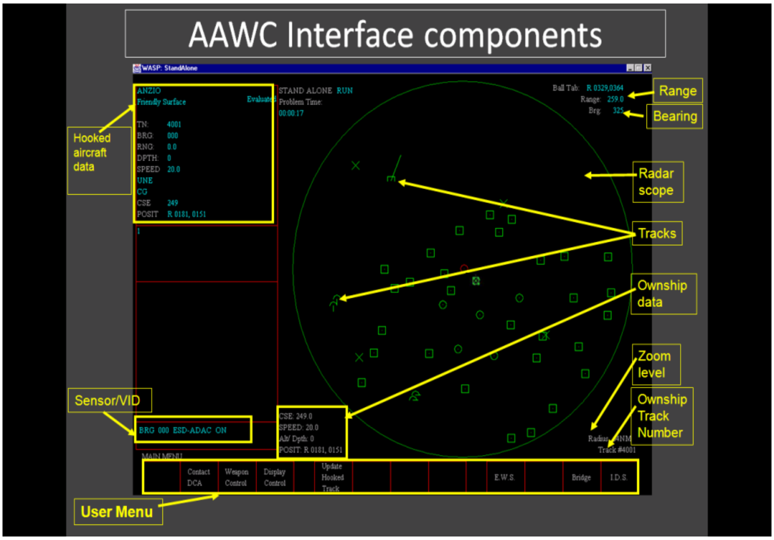


Fig. 3. AAWC interface components

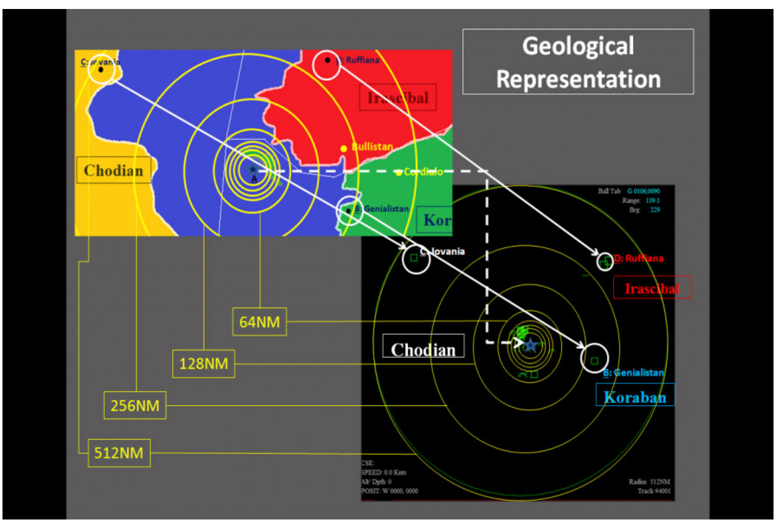


Fig. 4. Geographical relationship between Scenario Map and AAWC Interface

The user interface contains several components that aid the subject in his or her tasks. These provide dynamic information to participants in order to perform appropriate actions. In addition to visual feedback, auditory feedbacks corresponding to actions performed by participants are also provided from a sound library. AAWC test bed is used to simulate a complex command-and-control task environment in the laboratory setting to promote decision making under dynamic task conditions. Figure 4 shows the geographical representation used to create realistic experimental scenarios. Eight scenarios are developed to conduct the experiment. All events occurred in specific time sequences and were tied to Situation Awareness probes.

3.3 Procedure

Participants were asked to control the resources to perform tasks during the simulation. The experiment consists of two sessions – training session and actual trial session. Before the experiment, the participants were administered a questionnaire to solicit the experience on simulation task in the past. The participants participated in initial training session (Day 1) which lasted 60 minutes. During this session the participants were trained on the specific skills and provided with feedback on their task. Participants gained experience and understanding of the tasks based on the feedback provided by the instructor. The practice scenario lasted approximately 5 minutes with a task complexity that was easier than the actual trial. Based on the result of the pilot test, participants were ready to engage the actual trials after they executed the practice simulation for the third time. The participants after being trained underwent two experiment sessions (Day 2 and 3). Each subject experienced 8 scenarios. Each experiment session lasted approximately 90 minutes. During this session the participants were required to answer specific situation awareness questions that were designed for each scenario. Each scenario was designed to run 15 minutes long. Freeze occurred randomly between 10 and 15 minutes after the start of the scenario. Once the simulation was frozen, the participants answered situation awareness questions. Participants also required answering specific Pre-Meta and Post-Meta Probes. Based on the feedback type, participants experienced two different styles of Pre-Meta & Post-Meta Probes (different questions for each level of situation awareness for each scenario or one question for overall situation awareness for each scenario). Examples of Pre-Meta & Post-Meta Probes used for measurement are shown below:

- Pre-Meta & Post-Meta Probes for Skill based feedback and Control Group
  - Level1:“How well do you think you have detected the objects in your airspace?"
  - Level2:“"How well do you think you are aware of the current overall situation of your airspace?"
  - Level3:"How well do you think you are aware of where the overall situation of your airspace is heading?"

- Pre & Post Meta Probe for Overall feedback Group
  - Overall: "How well do you think you are aware of the objects and events in your airspace?"

We formed two experimental groups: skill- based feedback and overall feedback group. Each group was exposed to a different posttest feedback screen after participants had answered all probes. Participants who provided skill-based feedback monitored results of each level SA probes with each level of Pre-Meta and Post-Meta outcomes separately. Participants in the overall feedback group monitored overall result of all SA probes with their overall Pre-Meta and Pos-Meta rating. The control group did not receive any posttest feedback. In overall feedback trainees were allowed to view the feedback screen for a duration that was three times longer than the skill based feedback screen.

3.4 Performance Metrics

Based on the feedback type, participants experienced two different styles of Pre-Meta & Post-Meta Probes. Metacognitive behaviors are measured by the subjective ratings of Pre-Meta and Post-Meta Probes. Skill-based feedback group and control group had different questions for each level of situation awareness. Participants in these groups rated using % accuracy (1 to 100) scale. On the other hand, overall feedback group had one question for Pre-Meta and one for Post-Meta for each scenario (1 to 100) scale. The average response of all SA probes was used as a performance metric to assess participants' SA accuracy.

4 Results

To answer the research question, the influence of the posttest feedback in metacognitive monitoring process and the accuracy of situation awareness at dynamic control tasks, analysis of variance (ANOVA) was conducted. Figure 5 shows the mean and patterns of the participants' accuracy on the three groups.

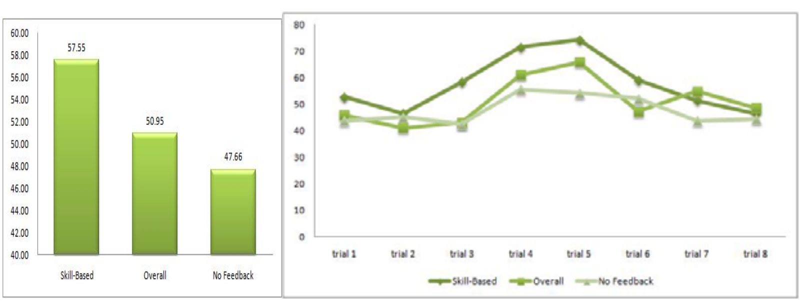


Fig. 5. Comparison of the results: SA accuracy

Table 1 shows the results of analysis of variance. It indicates SA accuracy of skill-based feedback group is significantly better than both overall and control group. However, there is no significant difference between overall feedback and control group on SA accuracy.

Table 1. One-Way ANOVA test of SA accuracy

	T-Value	*F-Value*	*P-Value*
Skill-Based vs. Overall Feedback	*2.72*	*7.39*	***0.007***
Skill-Based vs. No Feedback	*4.06*	*16.5*	***0.000***
Overall vs. No Feedback	*1.32*	*1.73*	*0.189*

Metacogntive behaviors were monitored using the subject ratings of Pre-Meta and Post-Meta probes. Table 2 explains the results of analysis of variance. It shows that subject's rating of Pre-Meta in skill-based feedback group was significantly higher than control group. For the Post-Meta, the rating of participants who were in the skill-based group had higher scores than other groups.

Table 2. One-Way ANOVA test of Pre and Post Meta

		T-Value	*F-Value*	*P-Value*
Skill-Based vs. Overall Feedback	*Pre*	*1.94*	*3.76*	*0.054*
	Post	*2.88*	*8.31*	***0.004***
Skill-Based vs. No Feedback	*Pre*	*3.05*	*9.3*	***0.003***
	Post	*2.18*	*4.74*	***0.03***
Overall vs. No Feedback	*Pre*	*1.07*	*1.14*	*0.287*
	Post	*-0.7*	*0.5*	*0.482*

5 Discussion

The present study compared the effects of the posttest feedback conditions in dynamic control tasks. Accuracy of situation awareness and subject rating of metacognitive behavior, Pre-Meta and Post-Meta, were monitored through the human-in-the-loop simulation. An important finding is the skill-based feedback group appears to outperform overall and no feedback group. In addition, three different patterns of metacognitive behavior were collected from the data.

Figure 6 indicated that the metacognitive monitoring activity has a strong impact on the operators' situation awareness performance due to the opportunity it provides to perform self error correction. This issue merits further attention. The preliminary findings of this study provide a better understanding of the impact of metacognitive training in the dynamic computer-based tasks.

Fig. 6. Metacognitive behavior (top: Skill-based, middle: Overall, bottom: No feedback)

References

1. Azevedo, R.: Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. Educational Psychologist 40, 199–209 (2005)
2. Flavell, J.: Metacognition and cognitive monitoring. American Psychologist 34, 906–911 (1979)
3. Siegler, R., Lin, X.: Self-explanations promote children's learning. Metacognition, Strategy Use, and Instruction, 85 (2009)
4. Afflerbach, P., Cho, B.: Determining and describing reading Strategies. Metacognition, Strategy Use, and Instruction, 201 (2009)
5. Jackson, N., Butterfield, E.: A conception of giftedness designed to promote research. Conceptions of Giftedness, 151–181 (1986)
6. Squire, L.: Mechanisms of memory. Science 232, 1612 (1986)
7. Sussan, D., Son, L.: The training of metacognitive monitoring in children. Columbia Undergraduate Science Journal 2 (2007)

8. Son, L.: Spacing One's Study: Evidence for a Metacognitive Control Strategy* 1. Journal of Experimental Psychology: Learning, Memory, and Cognition 30, 601–604 (2004)
9. Nelson, T., Narens, L.: Metamemory: A theoretical framework and new findings. The Psychology of Learning and Motivation: Advances in Research and Theory 26, 125–173 (1990)
10. Teong, S.: The effect of metacognitive training on mathematical word-problem solving. Journal of Computer Assisted Learning 19, 46–55 (2003)
11. Veenman, M., et al.: Initial inductive learning in a complex computer simulated environment: the role of metacognitive skills and intellectual ability. Computers in Human Behavior 18, 327–341 (2002)
12. Chi, M.: Knowledge structures and memory development (1978)
13. Endsley, M.: Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society 37, 32–64 (1995)
14. Endsley, M.: A survey of situation awareness requirements in air-to-air combat fighters. The International Journal of Aviation Psychology 3, 157–168 (1993)

The Effects of Personality Type in User-Centered Appraisal Systems

Zacharias Lekkas[1], Nikos Tsianos[1], Panagiotis Germanakos[2], Constantinos Mourlas[1], and George Samaras[2]

[1] Faculty of Communication and Media Studies, National & Kapodistrian University of Athens, 5 Stadiou Str, GR 105-62, Athens, Hellas
[2] Computer Science Department, University of Cyprus, CY-1678 Nicosia, Cyprus
zlekkas@gmail.com, {ntsianos,mourlas}@media.uoa.gr, {pgerman,cssamara}@cs.ucy.ac.cy

Abstract. The basic objective of this paper is to make an extensive reference of a combination of concepts and techniques coming from different research areas, Psychology and Web personalization technologies, both of which focus upon the user. It has been attempted to approach the theoretical considerations and technological parameters that can provide the most comprehensive user profile, under a common filtering element, supporting the provision of the most apt and optimized user-centered web-based result. It further underpins the significance of the comprehensive user profile that incorporates not only the traditional user characteristics, but other intrinsic values of the user such as user psychological preferences (affect, personality and emotional processing parameters). Eventually, this paper introduces our first experimental results that concern the concept of personality and its effect on decision making and problem solving user profiles.

Keywords: Personality, emotion, personalization, decision-making, problem solving, learning.

1 Introduction

Since 1994, the Internet has emerged as a fundamental information and communication medium that has generated extensive enthusiasm. The Internet has been adopted by the mass market much quicker than any other technology over the past century and is currently providing an electronic connection between progressive businesses and millions of customers and potential customers whose age, education, occupation, interest, and income demographics are excellent for sales. The explosive growth in the size and use of the WWW as well as the complicated nature of most Web structures result in orientation difficulties, as users often lose sight of the goal of their inquiry, look for stimulating rather than informative material, or even use the navigational features unwisely. As the e-Services sector is rapidly evolving, the need for such Web structures that satisfy the heterogeneous needs of its users is becoming more and more evident.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 388–396, 2011.
© Springer-Verlag Berlin Heidelberg 2011

To alleviate such navigational difficulties, researchers have put huge amounts of effort to identify the peculiarities of each user group and analyze and design methodologies and systems that could deliver up-to-date adaptive and personalized information, with regards to products or services. To date, there has not been a concrete definition of personalization. The many Web solutions offering personalization features meet an abstract common goal: to provide users with what they want or need without expecting them to ask for it explicitly [1]. Further consideration and analysis of parameters and contexts such as users' cognitive and mental capabilities, socio-psychological factors, emotional states and attention grabbing strategies should be extensively investigated. All these characteristics could affect the apt collection of users' customization requirements and along with the 'traditional' user characteristics (i.e. name, age, education, experience, profession etc.) constitute a comprehensive user profile that serves as the ground element of most of these systems offering in return the best adaptive environments to their preferences and demands.

Although one-to-one Web-based content provision may be a functionality of the distant future, user segmentation is a very valuable step in the right direction. User segmentation means that the user population is subdivided, into more or less homogeneous, mutually exclusive subsets of users who share common user profile characteristics enabling the possibility of providing them with a more personalized content. The subdivisions could be based on: Demographic characteristics (i.e. age, gender, urban or rural based, region); socio-economic characteristics (i.e. income, class, sector, channel access); psychographic characteristics (i.e. life style, values, sensitivity to new trends); individual physical and psychological characteristics (i.e. disabilities, attitude, personality and emotion).

Adaptivity is a particular functionality that alleviates navigational difficulties by distinguishing between interactions of different users within the information space [2]. The user population is not homogeneous, nor should be treated as such. To be able to deliver quality knowledge, systems should be tailored to the needs of individual users providing them personalized and adapted information based on their perceptions, reactions, and demands.

2 Elements of Web-Based Appraisal Systems

For many years people have been trying to measure differences between individuals. Over the course of time, a combination of developments in statistical know-how and the evolution of thought within psychology enabled the refinement of measures, and subsequently the assessment of more specific factors in the field of individual differences like different kinds of ability, affect and personality. This knowledge has been used in many areas within psychology and at the same time the advancement of technology has enabled the development of web-based appraisal systems that measure specific factors relevant to specific situations.

Web-based information systems are increasingly being used for learning and training applications. Computers are becoming better and more sophisticated every day. They can already perceive information related to user needs, preferences and characteristics [3, 4]. One possible implementation of a Web-based system's interface that can appraise human characteristics is through the use of a series of online tests and questionnaires that can assess the abilities and properties of the user [5]. E-assessment is the use of information technology for any assessment-related activity. Due to its obvious similarity to e-learning, the term e-assessment is becoming widely used as a generic term to describe the use of computers within the assessment process. E-assessment can be used to assess cognitive and affective abilities using e-testing software.

A web-based assessment system usually measures verbal and quantitative abilities but in order to acquire a more solid impression of the users' potential we include in addition some purely psychological constructs that help us measure terms like personality and affect. It does not measure a person's knowledge on specific fields of study and only minimal computer skills are needed. The Verbal section measures a user's ability to read and comprehend written material, to reason and evaluate arguments. Two types of multiple-choice questions are used in the verbal section, Reading Comprehension and Critical Reasoning. The Quantitative section measures a person's ability to reason quantitatively, solve quantitative problems, and interpret graphic data. Two types of questions are used in the quantitative section, Problem solving and Data sufficiency. Both types of question require basic knowledge of Arithmetic, Elementary algebra and commonly known concepts of geometry [6].

The use of psychometric data is considered a valuable tool for the evaluation of the individual. Psychometrics is an area of Psychology concerned with the systematic testing, measurement and assessment of aptitudes and personality. Psychometric assessments are used to complement less formal and more subjective methods to help teachers or managers reach more widely informed and objective judgements about people. We believe that by putting a special emphasis in the psychometric abilities of the user we will be able to examine a hidden but very important aspect of his behavior and performance.

3 Incorporating Cognitive and Affective Factors in the Personalization Process

The concepts of personality and affect underpin psychology's attempt to identify the unique character of individuals. The terms describe properties of behavior which concern the individual's typical ways of coping with life events [7]. An in-depth model that grasps the complexity of these underlying concepts is the first purpose of our research. Instead of selecting one area of implementation we are trying to combine these three levels of analyses (verbal, quantitative and psychological) and form a typology that will help us circle effectively the cognitive and affective mechanisms of the brain. In order to apply a purely psychological construct to a digital platform we adjust the various theories concerning cognition and emotion having in mind to make our model flexible and applicable to users' profiles, needs and preferences. The verbal and quantitative sections of the appraisal system are straightforward and cannot be significantly manipulated since they are easily quantifiable. In order to manipulate the parameters in the psychological section according to user characteristics, our research has to go through the stage of extracting quantified elements that represent deeper psychological and affective abilities. The latter cannot be directly used in a web environment, but a numerical equivalent can define a user characteristic. Apart from the standard personality questionnaire we developed a theory and a corresponding battery of questionnaires for the concept of Affect. Our psychological model of Affect has two base elements: Emotional Arousal is the capacity of a human being to sense and experience specific emotional situations. An effort to construct a model that predicts the role of specific emotions is beyond the scope of our research, due to the complexity and the numerous confounding variables that would make such an attempt rather impossible. We focus on arousal as an emotional mechanism and not on a number of basic emotions because emotional arousal can provide some indirect measurement of general emotional mechanisms since it manages a number of emotional factors like anxiety, boredom effects, anger etc. Our model would be problematic without a regulatory mechanism of affect. For this reason we constructed the measure of Emotion Regulation that is comprised from terms like emotional intelligence, self-efficacy, emotional experience and emotional expression. Emotion regulation is the way in which an individual is perceiving and controlling his emotions. Individuals attempt to influence which emotions they have, when they have them and how they experience and express them [8]. By combining the affective state of the individual with his regulatory mechanism we can reach into a conclusion of how affect influences his performance and the outcome of his behaviour.

4 The Concept of Personality

The personality theory that we follow is based on Eysenck's PEN model. The PEN model is comprised of three personality dimensions based on psychophysiology: Psychoticism, Extraversion and Neuroticism. These three dimensions are related to basic emotions and are considered to be superfactors that include lower-order factors such as sociability and positive affect. These lower order factors in turn include habits and behavioral patterns such as the components of sociability and impulsivity. Psychoticism is associated with the liability to have a psychotic episode or break with reality and also with aggression. Extraversion is related to social interest and positive affect, while Neuroticism is related with negative affect, stressors and pressure of many kinds. In this paper we present our first experimental procedure which investigates the connection between personality factors and decision making and problem solving styles. We consider important to clarify these basic and important concepts before examining deeper affective constructs. Personality is the doorway to emotional behavior and decision making and problem solving is an indirect way to make inferences to a person's learning pattern since learning includes continuous decision making and problem resolution [9].

5 Experimental Evaluation

5.1 Sampling and Procedure

The study was carried out within one week and the participants were all Greek citizens that live in Greece. All participants were of relatively young age studying or working at the time of administration. They could either participate in the experimental sessions that were held in the New Technologies laboratory in

University of Athens or fill in the questionnaires that could also be found online in the web page designed specifically for that purpose. They were all given a battery of questionnaires. A total of 247 questionnaires were completed and returned. 55 of them were half completed or had double answers and were omitted from the sample. Our final sample included 192 participants giving a completion rate of almost 80%.

Participants varied from the age of 18 to the age of 40, with a mean age of 27 and a standard deviation of 5. 73 respondents were male and 119 were female. Among other demographic characteristics that were examined were the profession and the computer experience level of each participant.

5.2 Questionnaires

The study used questionnaires to collect quantitative data. It included five measures, one each for personality, emotional arousal, emotion regulation, decision making styles and problem solving styles. Our first treatment involved the close examination of the personality questionnaire and its correlation with decision making and problem solving styles. To evaluate personality we used Eysenck's Personality Questionnaire, the revised short version (EPQR-S), a short version of EPQR with 48 items and three scales (Extraversion, Neuroticism, Psychoticism), suitable for web environments because it can be administered fast and it is accurate enough in predicting user behaviour [10]. For Decision Making we used the General Decision-Making Style Inventory (DMSI) by Scott and Bruce [11] which includes 25 items and 5 scales (Spontaneous, Dependent, Rational, Avoidant, Intuitive) and for Problem Solving the Problem Solving Styles Questionnaire (PSSQ) by Parker with 20 items and four scales (Sensing, Intuitive, Feeling, Thinking).

5.3 Design

Internal consistency was assessed by computing Cronbach alphas for the three measures. Although there are no standard guidelines available on appropriate magnitude for the coefficient, in practice, an alpha greater than 0.60 is considered reasonable in psychological research [12]. After the inspection of the alpha coefficients, we performed descriptive statistics for the study sample as a whole and for the particular scales under investigation to examine the sample's suitability. Since our sample was normally distributed with variance of suitable proportions we continued our statistical analysis with the use of the statistical package SPSS. The statistical analysis used to perform this study was mainly one-way Analysis of Variance (ANOVA). Our research hypothesis stated that personality factors will have an effect on the participant's style of action. More specifically, participants that score high in Neuroticism, Extraversion and Psychoticism scales will have a tendency towards more emotional and less rational styles.

6 Results

For the purposes of the experiment, Analyses of Variance (ANOVA) were performed in order to indicate the relationships between the variables of the study. Table 1 presents the main findings between the scale of neuroticism and the scales of the DMSI and PSSQ. The analyses indicated that neuroticism correlated highly with the

Table 1. Statistical Significance between the Neuroticism scale and Decision-Making and Problem-Solving Styles

Constuct	F	Sig.
DM-Spontaneous	4.422	.037*
DM-Rational	4.888	.028*
DM-Avoidant	5.319	.022*
PS-Feeling	9.579	.002**
PS-Thinking	5.005	.026*

*p<0.05
**p<0.01

spontaneous, rational and avoidant styles of the decision making questionnaire and the feeling and thinking styles of the problem solving questionnaire.

A person highly neurotic is typically anxious, tense and moody. He can get emotional easily and therefore is reasonable to react in a spontaneous and not thoughtful way in occasions or with an inhibition of action in others. His pattern of behavior is tense as his character and is subjective to strong feelings. On the other hand a stable person is more rational and more methodical in his behavior. In figure (graph) 1 we can clearly see that neurotics have a higher mean in spontaneous, avoidant and feeling scores while stable participants have higher scores in rational and thinking styles.

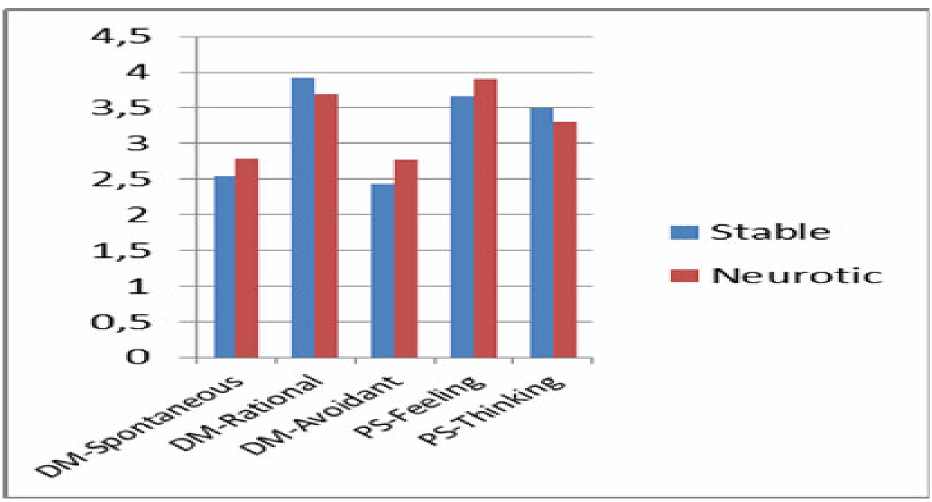


Fig. 1. Means of neurotic and stable participants in the statistically significant styles. Neurotic participants have higher means in the more "emotional" styles of spontaneous, avoidant and feeling while stable participants score higher in the "logical" ones such us rational and thinking.

The Psychoticism scale includes traits such as egocentric, dogmatic, tough-minded and aggressive. People that score highly in this scale tend to react on their own. In table 2 can be seen the significant differences between people with high psychoticism score and people with low.

Table 2. Statistical Significance between the Psychoticism scale and Decision-Making and Problem-Solving Styles

Constuct	F	Sig.
DM-Spontaneous	4.719	.031*
PS-Intuitive	4.905	.028*

*p<0.05

The extraversion construct includes traits such as sociable, sensation-seeking, impulsive, expressive and active. On the contrary, the other side of the scale, the introvert is more reflective and more centered in the inner energy of his self. Table 3 shows once again the significant differences with styles.

Table 3. Statistical Significance between the Extraversion scale and Decision-Making and Problem-Solving Styles

Constuct	F	Sig.
DM-Intuitive	6.733	.010*
PS-Sensing	9.543	.002**
PS-Feeling	4.345	.039*

*p<0.05
**p<0.01

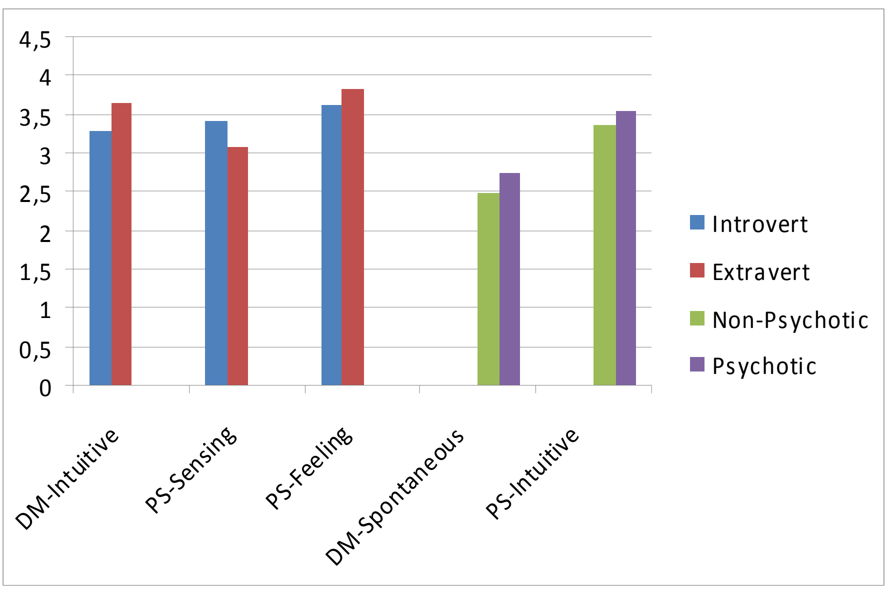


Fig. 2. Means of psychotic and extravert participants in the statistically significant styles

In figure (Graph) 2 we can see that those that score high in psychoticism are more spontaneous and intuitive probably because they are not co-operative and often react emotionally. In the same graph it is depicted that extraverts, as predicted, show higher scores in intuitive and feeling styles because they will not examine the facts closely, while introverts score higher in the sensing category as it is reasonable because they tend to put an emphasis on the detail.

7 Discussion

It may come as no surprise that personality factors are important in the decision and problem solving process. Personality traits such as Neuroticism, Psychoticism and Extraversion are comprised from characteristics that people often exhibit in their decision making. The analyses indicated that participants high in the neuroticism scale were more emotional than stable participants low on neuroticism. Stable participants across the various conditions proved to be more logical and straightforward. As it is shown, extraverts are more likely to experience emotions while individuals with high neuroticism and psychoticism tend to experience negative emotions especially anxiety.

Apart from the standard personality questionnaire we developed a theory and a corresponding battery of questionnaires for the concept of Affect [13]. The next step of our research is to combine these findings with the purely affective elements of our model. It has been argued that positive affect increases motivation, attention, pleasantness, participation and engagement, while negative affect is highly involved with boredom, fear, anger, displeasure and distraction.

By combining the personality style and the affective state of the individual with his regulatory mechanism we can reach into a conclusion of how affect influences his performance and the outcome of his behavior. At the same time our level of implementation after analysing our findings in decision making and problem solving preferences, will concentrate directly on the user learning process. We have already developed a web-based system based on learning performance evaluation for the testing of the various instruments that we have incorporated in our model [14]. The cognitive elements are more straightforward since they are easier to measure and easier to quantify and we have already reached a level in which we can make inferences about how users with different cognitive abilities and preferences can be aided or guided through a personalized web interface [15]. The final step to complete the implementation of our model is to add the affective elements and to investigate the inner and deeper relations that exist between them. Personality type is the fundamental construct since personality research is already established and developed to a great extent. Our next task is to examine our findings in combination with the constructs of task-specific anxiety and emotion regulation.

Acknowledgements. This work is co-funded by the EU project CONET (INFSO-ICT-224053) and by the project smarTag (within the framework of University of Cyprus internal funded projects).

References

1. Mulvenna, M.D., Anand, S.S., Buchner, A.G.: Personalization on the net using Web mining. Communications of the ACM 43(8), 123–125 (2002)
2. Eklund, J., Sinclair, K.: An empirical appraisal of the effectiveness of adaptive interfaces of instructional systems. Educational Technology and Society 3(4) (2000)
3. Cingil, I., Dogac, A., Azgin, A.: A broader approach to personalization. Communications of the ACM 43(8) (2000)
4. Kim, W.: Personalization: Definition, Status, and Challenges Ahead, Zurich. Chair of Software Engineering JOT 1(1) (2002)
5. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
6. The e-Assessment Association, `http://www.e-assessment.com`
7. Lewis, M., Haviland-Jones, J.M.: Handbook of emotions, 2nd edn. The Guildford Press, New York (2004)
8. Bechara, A., Damasio, H., Damasio, A.R.: Emotion, decision-making, and the orbitofrontal cortex. Cerebral Cortex 10, 295–307 (2000)
9. Kline, P.: Personality. The Psychometric View. Routledge, London (1993)
10. Francis, L.J., Philipchalk, R., Brown, L.B.: The comparability of the short form EPQR with the EPQ among students in England, the U.S.A., Canada and Australia. Personality and Individual Differences 12, 1129–1132 (1991)
11. Scott, S.G., Bruce, R.A.: Decision-making style: the development and assessment of a new measure. Educational and Psychological Measurement 55(5), 818–831 (1995)
12. Kline, P.: Handbook of Psychological Testing. Routledge, London (2000)
13. Lekkas, Z., Tsianos, N., Germanakos, P., Mourlas, C.: Integrating Cognitive and Emotional Parameters into Designing Adaptive Hypermedia Environments. In: Proceedings of the Second European Cognitive Science Conference (EuroCogSci 2007), pp. 705–709 (2007)
14. Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., Samaras, G.: Capturing Essential Intrinsic User Behaviour Values for the Design of Comprehensive Web-based Personalized Environments. Computers in Human Behavior Journal, Special Issue on Integration of Human Factors in Networked Computing (2007)
15. Tsianos, N., Lekkas, Z., Germanakos, P., Mourlas, C., Samaras, G.: User-centered Profiling on the basis of Cognitive and Emotional Characteristics: An Empirical Study. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) AH 2008. LNCS, vol. 5149, pp. 214–223. Springer, Heidelberg (2008)

Measuring Corrective Reaction Time with the Intermittent Illumination Model

Jui-Feng Lin[1], Colin G. Drury[2], Chin-Mei Chou[1], Yu-De Lin[1], and Yi-Quan Lin[1]

[1] Department of Industrial Engineering & Management, Yuan Ze University, Chung-Li, Taiwan 320
[2] Department of Industrial & Systems Engineering, State University of New York at Buffalo, Buffalo, New York 14260, USA

Abstract. The main objective of this research was to develop a methodology based on Drury's [1] intermittent illumination model to directly measure individuals' corrective reaction times, without the two methodology-related issues of Lin & Drury [2]. Six highly-practiced participants performed self-paced circular tracking tasks by using a modified LED monitor in a darkened room. While performing movements, the monitor's backlight was intermittently turned on and off to generate five values of expected delay of visual feedback. Expected delay and measured speed were used with the intermittent illumination model to calculate individuals' corrective reaction times. The results of showed that the model fitted the data very well, accounts for at least 94.6 % of the variance. The mean corrective reaction time was 273 milliseconds and ranged from 170 to 460 milliseconds for individual participants. While previous studies only reported group means, this was the first study to report individuals' corrective reaction times.

Keywords: corrective reaction time, intermittent illumination model, hand-control movement, tracking movement.

1 Introduction

Corrective reaction time is an important component of the study and modeling hand-control movements. Due to the psychological refractory period [3], the utility of visual feedback on movement accuracy is intermittent, instead of continuous. Hence, a hand-control movement, such as pointing a fingertip to click a button on a touch screen, might be consist of more than one ballistic movement, each correcting the distance remaining between the fingertip and the button. The shortest time between the end of one ballistic movement and the start of the next is called the "corrective reaction time" during which the brain receives visual feedback, programs a movement impulse, and sends the impulse to the controlled limbs. To better understand our motor control system and to predict the speed-accuracy tradeoff relationships while performing hand-control movements, it is necessary to measure the corrective reaction time effectively, particularly if we are to understand individual differences.

Early findings on corrective reaction time come mainly from the studies of aiming movement. Woodworth's [4] pioneer work showed that aiming movements made at a

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 397–405, 2011.
© Springer-Verlag Berlin Heidelberg 2011

rate of 140 times/minute or greater were equally accurate no matter whether the visual feedback of ongoing movements was available or not. This led him to conclude that the time required to process visual feedback for movement control is about 450 milliseconds. This finding was further supported by Vince [5] who replicated Woodworth's experiments. However, Keele & Posner [6] argued that the 450 milliseconds as the corrective reaction time was overestimated. The reason was because that the experimental tasks conducted by Woodworth [4] and Vince [5] were reciprocal movements in which the measured time included the time spent on reversing the movement direction after targets were reached. To solve the issue, instead of reciprocal movements, Keele & Posner [6] utilized discrete aiming movements as their experimental tasks and found out that the time required for visual feedback loop to operate was somewhere between 190 and 260 milliseconds. Later, Beggs & Howarth [7] utilized a different experimental paradigm and obtained a result close to Keele & Posner's [6]. Their participants performed sagittal aiming movements that started from a position 10 cm above their shoulder and ended at a cross target vertically attached on a wall placed in front of them. When they performed the movements, the initial part of the movement trajectory was illuminated, but the room lights were extinguished when the pen reached a particular distance from the target. Their results showed that movement accuracy was not improved once the availability of visual information was shorter than 290 milliseconds before the target was reached. All the experimental methods mentioned above estimated corrective reaction time by statistically comparing the effects of visual feedback on movement accuracy. Although the methods were rough and the individual differences were ignored, they gave an approximate range from 190 to 290 milliseconds of corrective reaction time.

Beside the studies related to aiming movements, Drury's [1] intermittent illumination model has theoretical support and can also predict corrective reaction time. In the experiment by Tsao & Drury [8], participants were asked to draw lines within circular courses as fast as they could while maintaining specified accuracy. When they performed the drawing tasks, the room lights were manipulated to provide intermittent illumination comprising different periods of dark and light. To model the effect of intermittent illuminations on movement speeds, Drury [1] integrated the models by Drury [9] and Howarth, Beggs, & Bowden [10] and then theoretically developed the intermittent illumination model (Equation 1).

$$\frac{1}{c} = K \times \sigma_\theta \times (t_r + \text{Expected Delay}) \tag{1}$$

$$\text{Expected Delay} = \frac{d^2}{2(l + d)} \tag{2}$$

where, c is the controllability [9], K is a constant [10], σ_θ is the angular accuracy [10], l is the light period, d is the dark period, and t_r is the corrective reaction time. The model predicts the linear relationship between the inverse of the controllability and the visual feedback cycle time, $t_r + \frac{d^2}{[2(l+d)]}$, which represents the sum of the corrective reaction time and the Expected Delay manipulated by the intermittent illumination of *l* and *d*. The results from Tsao & Drury's [8] experimental data showed that Equation (1) explained over 90% of the variance in the slopes of the speed/width regression for a variety of dark and light intervals. Furthermore, the model gave an estimate for the mean corrective reaction time as 238 milliseconds. As Keele & Posner's [6] and Beggs & Howarth [7], Drury [1] did not report individual differences.

Recently, Lin & Drury [2] redesigned Tsao & Drury's [8] experimental method and attempted to precisely measure individuals' corrective reaction times. Instead of drawing circles on white paper, their participants used a drawing tablet to control the screen cursor to draw circles within circular courses shown on the screen. To generate the expected delay, instead of controlling the intermittent illumination, the visual information of the cursor was intermittently displayed. While performing movements, the cursor started to blink according to predetermined light/dark periods. However, their measurement of the corrective reaction times did not match the results reported in the literature. The two potential experimental design issues were reported as (1) inappropriate manipulations of the expected delay and (2) indirect visual feedback of control movements. Hence, the main purposes of this study were to deal with the two issues mentioned by Lin & Drury [2] and thus develop a standard methodology to measure individuals' corrective reaction times using the intermittent illumination model.

2 Method

2.1 Participants

Six undergraduate students, two male and four female, aged from 19 to 20 years, were recruited to participate in this study. They were all right-handed with normal or corrected-to-normal vision.

2.2 Apparatus

A personal computer (PC) with a 19" (48.3 cm) modified LED monitor was used. To deal with the indirect visual feedback issue mentioned by Lin & Drury [2], the monitor was attached with a touch panel so that the circular tasks could be performed directly on the monitor. Further, the monitor was modified to a condition where its backlight could be turned on and off intermittently to generate intermittent visual feedback of the tracking circular courses, drawing stylus and the control limbs.

The experimental program was developed similar to that used in Lin & Drury [2]. The PC ran Visual Basic using this program to display the circular courses, generate the intermittent illuminations and measure movement speeds.

2.3 Experimental Setting and Procedures

The experiments were conducted in a darkened room and the only illumination source was provided by the modified LED monitor. As shown in Fig. 1, while performing experimental tasks, participants sat on a chair. The monitor was placed on an adjustable stool placed between participants' legs. Predetermined circular courses with a mean radius of 200 pixels (53.2 mm) and different course widths were displayed randomly on the screen. To perform tracking movements, participants pressed down the stylus cursor on the start point placed at the top location of the courses. Once the cursor was moved away from the start point, the start point disappeared and the backlight of the monitor started to blink according to predetermined appearing / disappearing periods (see Table 1). For each circular course, participants needed to draw three continuous circles in clockwise direction. They were asked to draw as quickly as possible, but without moving outside the courses. If the cursor was moved outside the courses, that movement was considered as a failure trial. They had to repeat that course until it was successfully completed. To familiarize the participants with the tasks and to optimally utilize the intermittent visual feedback, they had extensive practice in all tasks, six to ten hours, in the week before formal measurements. Before each formal measurement, they performed a five-trial training module to complete dark adaptation and to re-familiarize with each task.

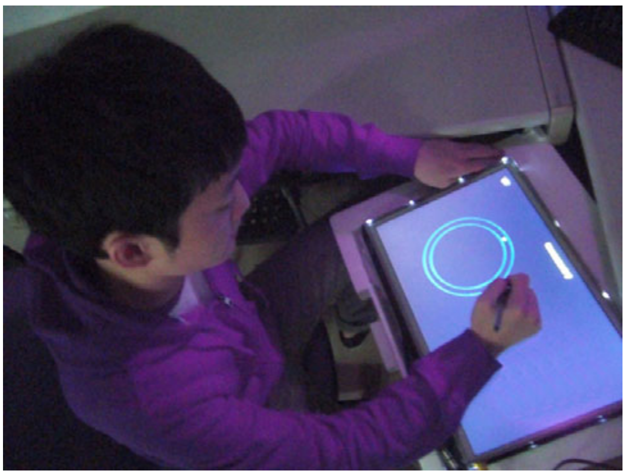


Fig. 1. Executions of circular tracking movement

2.4 Experimental Variables

The independent variables were: Course Width and Expected Delay. The four values of Course Width were 30、35、40 and 45 pixels (1 pixel ≅ 0.266 mm). The five values of Expected Delay with their combinations of light and dark durations determined by Equation 1 are listed in Table 1. The experimental combinations in this experiment were replicated twice, resulting in a total of 40 trials. All the trials were presented to each participant in a random order, taking about an hour to complete. Each participant needed to complete eight formal measurements arranged over two consecutive weeks.

The only dependent variable was speed. In a three-circle tracking movements, speed was measured per quarter circle. To ensure that movements were measured with consistent speed, the first circle and the last quarter circle of the three circles were not measured. Hence, seven speed values were measured in a three-circle trial.

Table 1. Combinations of dark and light durations of the five values of expected delay

Expected Delay	Dark Period (l)	Light Period(d)
0 ms	0 ms	∞ ms
150 ms	58 ms	350 ms
300 ms	54 ms	650 ms
450 ms	53 ms	950 ms
600 ms	52 ms	1250 ms

3 Results

3.1 Analysis of Variance

Analysis of variance was performed on the speeds, using a mixed model with Width and Expected Delay as fixed effects and Participant as random, in which the two-way and three-way interaction effects among all the effects were analyzed. Despite of Participants (significant at $p = 0.006$), all the other main effects and all the interaction effects were highly significant ($p < 0.001$). Note that because of large numbers of measurement (total df = 13439), these significant effects were not practical. The main effect of Participant (see Fig. 3) showed that the participants performed the movement at different speeds. The main effects and the two-way interaction effect of Width and Expected Delay are shown in Fig. 2, where (1) the increase of Course Width resulted in increased speed, (2) the increase of Expected Delay resulted in decreased speed and (3) these rates of decrease of speed were especially large when Expected Delay increased from 0.0 to 0.15 s, representing the interaction effect of Width × Expected Delay.

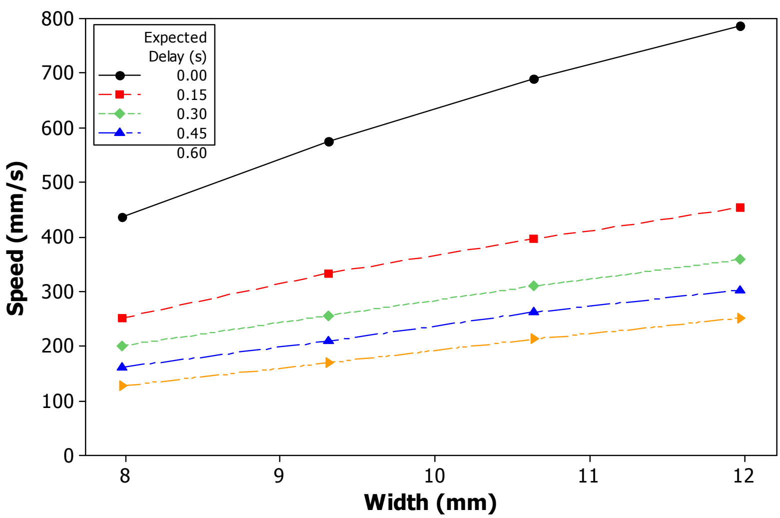


Fig. 2. The effects of Width and Expected Delay on speeds

Since the main effects of Course Width and Expected Delay were found significant, the application of the intermittent illumination model could be tested. The calculated mean 1/*c* values across all participants and for individual participants were regressed on to Expected Delay (Equation 2) to give the intercept (a), slope ($K \times \sigma_\theta$), r^2, and corrective reaction time (t_r). All of the regressions explained at least 97.2 % of the variance and were significant at $p \leq 0.001$, while the overall regression accounted for 99.4% of variance and was significant at $p < 0.001.$ The regressions specified to the overall participants' measurements and individuals' measurements are shown in Fig. 3. The calculated corrective reaction time was 273 milliseconds for overall participant, and it ranged from 170 to 460 milliseconds for individual participants.

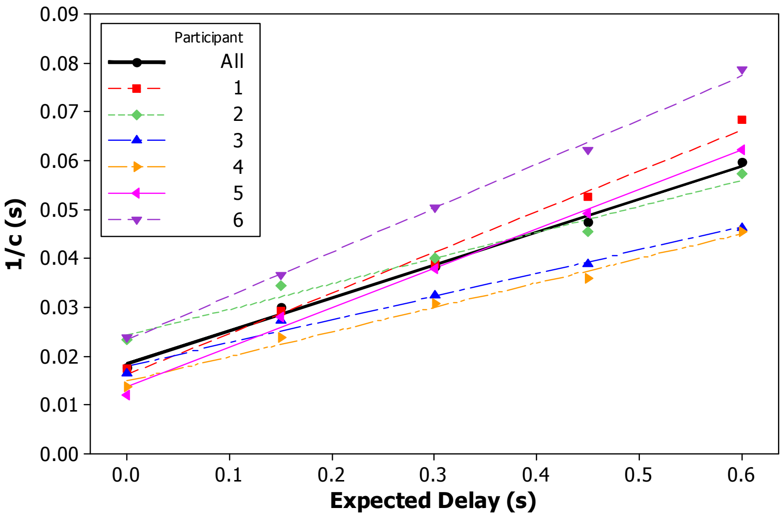


Fig. 3. Relationships between reciprocal controllability and expected delay for individual participants and mean values

3.2 Corrective Reaction Time

To analyze individual differences, the corrective reaction times calculated from every measurement were analyzed by a one-way ANOVA, followed by a Tukey means comparison test. The results showed that there were significant individual differences ($F_{5,42} = 31.85$, $p < 0.001$). Tukey's grouping information grouped participants 2 and 3; participants 4 and 6; and participants 1, 5 and 6. Graphic representation of individual differences is shown in Fig. 4.

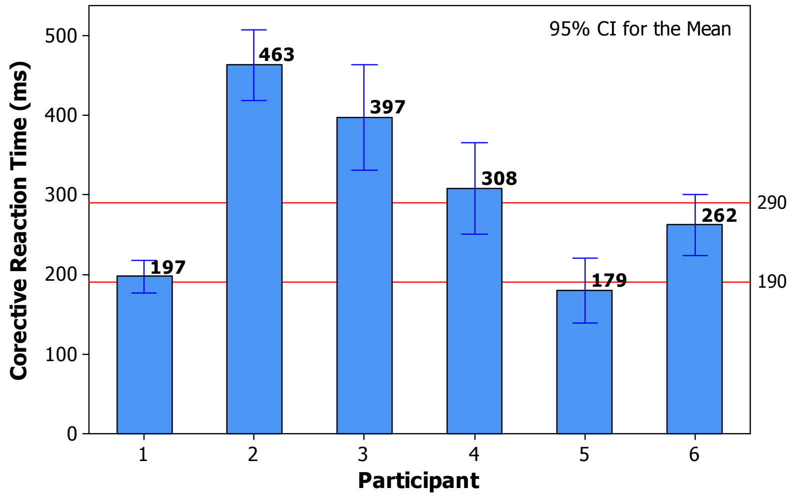


Fig. 4. Mean corrective reaction times for individuals

4 Discussion

As predicted by Drury's [9] and Drury's [1] two models, increased course widths and decreased expected delay value resulted in increased movement speed. Movement speeds measured under different expected delay values increased gradually as course width increased. Moreover, movement speeds measured under different course widths decreased linearly according to the increase of expected delay. However, the rate of decrease was greater when the expected delay value increased from 0 to 0.15 s compared to other changes. There was clearly a large difference between no delay and some delay.

By overcoming the two methodology-related issues of Lin & Drury [2], this study successfully measured individuals' correction reaction time. To make sure that participants kept moving in the dark duration of the intermittent illumination, in this study, the expected delay values were manipulated by keeping dark durations were kept close (52 - 58 milliseconds) and varying light durations (350 - 1250 milliseconds). Note that in Lin & Drury [2], light durations were kept close (292 - 354 milliseconds) and dark durations were varied (500 – 1750 milliseconds). Further, to make the visual feedback of the movements was directly from the controlled stylus, the participants in this study drew the tracking tasks on a modified LED monitor. This experimental design was similar to that of Tsao & Drury [8], where the participants used a pen to drew circles on papers. With these two experimental modifications and by highly practicing participants before formal measurements, we showed the good fitting of the intermittent illumination model [1]. The model accounted for 99.4 % variance of the overall measurements. For individuals' eight measurements, the model accounted for an average of 98.6 % of the variance and at least accounts for 97.2 % of the variance.

While only group means were reported in the literature, this was the first study to report individual corrective reaction times. The calculated mean corrective reaction time for six participants in this study was 273 milliseconds, which in line with the previous studies [i.e., 1, 6, 7] that reported a range of 190-290 milliseconds as group mean corrective reaction time. Moreover, this study reported individuals' corrective reaction times that ranged from 170 to 460 milliseconds. While individual differences were studied, this study reported a wider range of the corrective reaction time. because of these individual differences, future research, e.g. the general model proposed by Lin, Drury, Karwan, & Paquet [11], can proceed with more confidence by modeling individuals as well as means across groups.

5 Conclusions

This study was another test of applying Drury's [1] intermittent illumination model to effectively measure individuals' corrective reaction times. To deal with the two issues mentioned by Lin & Drury [2], this study utilized a modified LED monitor and tested with different combinations of dark periods and light periods to generate the expected delay values for the model. The results showed that the model fitted the relationship between the reciprocal controllability and the expected delay very well. Furthermore, a mean 273 milliseconds corrective reaction time was obtained for overall participants. Individual differences in corrective reaction time ranged from 170 to 460 milliseconds. The measured range of corrective reaction time was wider than that reported in the literature, we considered that the results were more convincing as the regressions were highly linear and the resulting error variances in corrective reaction time were thus small. Future research will be able to use individual differences with confidence in modeling movements where corrective reaction time is a factor.

Acknowledgments. We would like to acknowledge the grants support from Taiwan National Science Council (NSC98-2218-E-155-11 and NSC 99-2221-E-155 -066).

References

1. Drury, C.G.: A model for movements under intermittent illumination. Ergonomics 37(7), 1245–1251 (1994)
2. Lin, J.-F., Drury, C.: An application of the intermittent illumination model for measuring individual's corrective reaction time. In: The 3rd International Conference on Applied Human Factors and Ergonomics, July 17-20. Taylor & Francis, Miami (2010)
3. Welford, A.T.: The "psychological refractory period" and the timing of high-speed performance - a review and a theory. British Journal of Psychology 43, 2–19 (1952)
4. Woodworth, R.S.: The accuracy of voluntary movement. The Psychological Review Monographs Supplement 3(13), 1–114 (1899)
5. Vince, M.A.: Corrective movements in a pursuit task. Quarterly Journal of Experimental Psychology 1, 85–103 (1948)
6. Keele, S.W., Posner, M.I.: Processing of visual feedback in rapid movements. Journal of Experimental Psychology 77(1), 155–158 (1968)

7. Beggs, W.D.A., Howarth, C.I.: Movement control in a repetitive motor task. Nature 225, 752–753 (1970)
8. Tsao, Y.C., Drury, C.G.: Self-paced tracking under intermittent illumination. Unpublished paper, Department of Industrial Engineering, State University of New York at Buffalo (1975)
9. Drury, C.G.: Movements with lateral constraint. Ergonomics 14(2), 293–305 (1971)
10. Howarth, C.I., Beggs, W.D.A., Bowden, J.M.: The relationship between speed and accuracy of movement aimed at a target. Acta Psychologica 35, 207–218 (1971)
11. Lin, J.-F., Drury, C., Karwan, M., Paquet, V.: A general model that accounts for Fitts' law and Drury's model. In: Proceedings of the 17th Congress of the International Ergonomics Association, Beijing, China, August 9-14 (2009)

Psychometric Evaluation with Brain-Computer Interface

Paolo Perego[1], Anna Carla Turconi[2], Chiara Gagliardi[2], and Giuseppe Andreoni[1]

[1] Politecnico di Milano, INDACO dept., via Durando 38/a, 20133 Milan, Italy
[2] IRCCS E. Medea, via don Luigi Monza 20, 23842, Bosisio Parini (LC), Italy

Abstract. Brain-Computer Interfaces (BCIs) are systems which can provide people affected by severe neuromuscular diseases with a new and easy way to communicate with the world. The aim of this study is to use a new administration method based on BCI technology to assess cognitive ability in disabled people. The protocol was applied on 15 subjects who can't or have difficult using traditional paper based test. The method, based on a SSVEP BCI system, was previously validated on 20 healthy subjects. The previous validation and the test results on disabled patient show the reliability of BCI in administering cognitive test; BCI doesn't affect results but adds further data that can be used to analyze other cognitive tasks in addition to those measured by the test.

Keywords: Brain Computer Interface (BCI), Steady State Visual Evoked Potential (SSVEP), psychometric assessment, Raven test.

1 Introduction

Many neuromuscular diseases and trauma (eg. ALS – Amyotrophic Lateral Sclerosis, CP – Cerebral Palsy, Brainstem stroke) can seriously disrupt the neuromuscular channels through which the brain communicates with and controls its external environment. This can cause motor impairments and, in some cases, complete paralysis (locked-in syndrome). Due to the loss of control of all voluntary muscles, including eye movements and respiration, the patients may be completely locked-in into their bodies, bedridden and unable to communicate in any way. In the last decades modern life-support technologies can allow most individuals, even locked-in patient, to live a relative long life. In these cases, the ability to communicate and interact with the external world is a must that can significantly improve their life quality. When other assistive technologies can't be used due to the lack in movement control (eye-tracker needs eye movement control) the last option is a new non-muscular communication channel that allows for a direct brain–computer interface (BCI) connection to transmit messages and commands to the external world [1]. The most used BCI system [1] detects changes in electro-encephalographic signal and, through a real-time processing, is able to extract commands which can be used, with a personal computer, in order to communicate both with other people and devices.

Since first BCI experiment in 1970s [2] until now, BCI reliability has grown rapidly allowing the realization of BCI systems which can be used in different area (BCI

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 406–413, 2011.
© Springer-Verlag Berlin Heidelberg 2011

for home automation [3], BCI for painting [4], BCI for communication [5]) but only few clinical experiences were carried out. Iversen et al [6] use SCP (Slow Corical Potential) based BCI to assess cognitive function in ALS patient through a not validated test and assuming if subjects can control a BCI system with two-choice task, they should answer question related to their cognitive skills. They didn't consider the possibility that the BCI system may need proper cognitive skills that can affect the test response.

Our administration protocol was validated on twenty healthy subjects and demonstrated SSVEP based BCI doesn't affect the test result. In this work we use this validated administration method with an already validated SSVEP base BCI system [7] to assess cognitive ability in people with motor impairments who either could not be tested.

2 Materials and Methods

Proposing a new administration method based on BCI technology for widely used clinical test is a must so that the presence of a validation with large control group allows verifying the results. For this reason we prefer to divide our study into two parts. In the first part we designed the protocol and validated it on 20 healthy subject demonstrating how the BCI cognitive weight doesn't affect the test result. In the second part, described in this paper, we recruit 26 subjects with different neurological diseases (cerebral palsy, dystrophies and paresis) to test the system.

2.1 The Administration Protocol

The administration method is composed by two main parts: the protocol and the software. The protocol consists of all the procedures which are necessary to complete the test. Figure 1 shows the flow chart of the entire administration protocol.

This protocol proceed by steps and the subject go to the next step if he/she is able to respond correctly to the four exclusion criteria.

The protocol starts with a checkup session aimed at identifying the BCI configuration parameters and then configure, train and validate the BCI classifier. Before this checkup session, the protocol provided 15 minutes for the mounting of the electrodes.

The first exclusion criteria excludes uncooperative subjects (like dystonic subject with involuntary head movement which affect EEG recording) or montage with high electrode impedances (>20KΩ). In the checkup session we find:

- *Screening phase* to check BCI suitability;
- *Training phase* to set up BCI classifier;

This part of the protocol is based on a previous experience [7]. The entire session during about 10 minutes and can be repeated to improve the performance of the classifier. The performance is in fact related with user ability to interact with the BCI and can increase with training. If the subject doesn't show, or had a weak SSVEP pattern, the classifier is not able to extract the right command and the subject is classified as being unable to use the BCI system and is excluded from the test as shown in figure 1 by the second and third exclusion criteria.

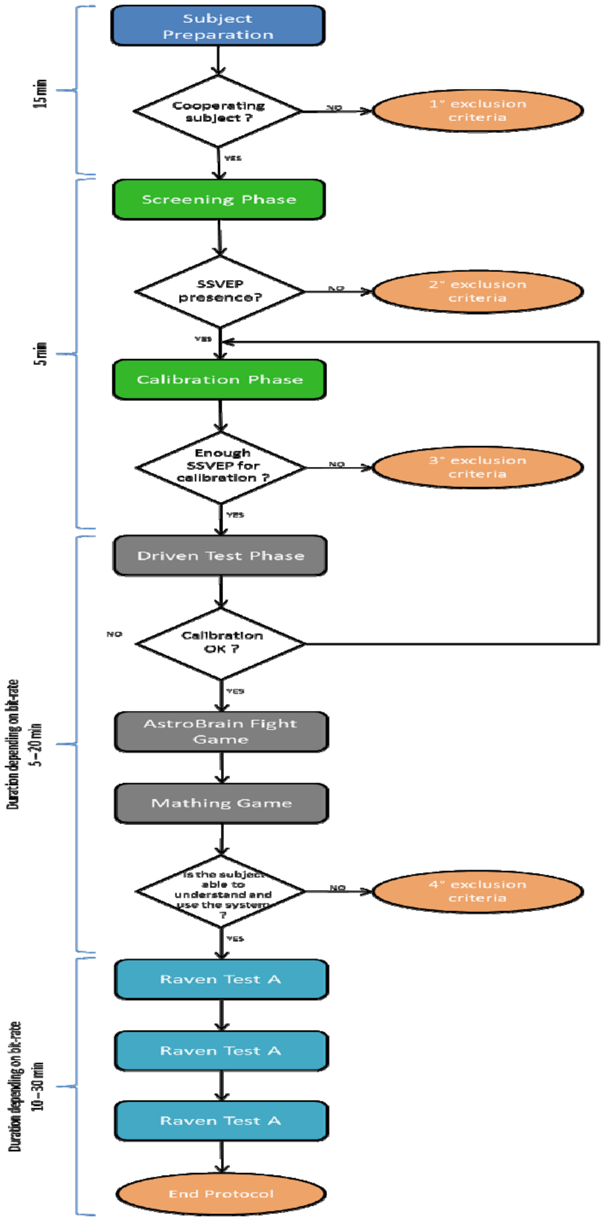


Fig. 1. Administration protocol. The entire test can be performed in a single or two separated sessions.

The protocol proceeds with the Learning Session. It is composed by:

- *Drive Test* to verify and validate the BCI system performance;
- *AstroBrain Fight Game* to train the subject using the BCI system;
- *Matching Game* to train the subject using the selection method that will be used with BCI for the cognitive test.

This session aimed to verifying the presence of specific visual perceptual competencies underpinning tasks like scanning or matching abilities. If the subject is not able to use BCI system he/she is excluded from the test (fourth exclusion criteria). Moreover, in this session the subject learns how to use the BCI system. After this session a short checkup can be repeated in order to improve the quality of the classifier: the subject is now able to use BCI and a recalibration could improve BCI reliability and performance.

The last session is the Testing Session. It lasts from 10 to 30 minutes (depending on the skill of the subject). In this session the subject performs the psychometric test. We use the Raven's Coloured Progressive Matrices (RCPM) to test general non-verbal intelligence. RCPM is the shorter and simpler form of the Raven's Progressive Matrices Test. It consists of 36 non-representational coloured pictures incompleted in the bottom right corner (Figure 2), organized in three sets of 12 diagrammatic puzzles; the subject is given six alternatives from which he/she has to select the one that best completes the given pattern. We chose this test because it is a standardized and already validated test, there are published normative data, it is a non-verbal test (avoiding the influence of language and limited language ability on the result) already widely used in clinic and, above all, it is easy and fast to administer both by paper or computer.

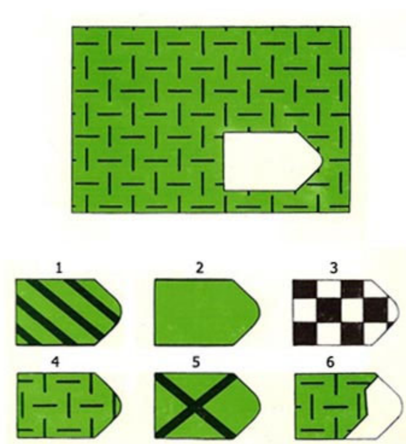


Fig. 2. Example of Coloured Raven Progressive Matrix

2.2 The BCI System

In this work we use our already validated SSVEP-based BCI system [7]. It is an user driven dependent BCI which allows the subject to send four different commands to the user interface. SSVEP BCIs are the most used and efficient system [8]; this is the reason why we chose this particular BCI system. However the software has been developed independently from the BCI system and P300, SCP and Motor Imagery could be also used in the future.

The BCI is composed by two modules, one dedicated to the recording and processing of the 8 channels EEG (placed according to the 10-20 system: O1,Oz,O2,P3,Pz,P4,T5,T6) and the other to the user interface. These modules run on two separated PC in order to leave the subject in a separated room without external stimuli or noise which can affect the test results. The first module is the Hardware Interface Software (HIM)[9] which is the program dedicated to the acquisition, processing and visualization of the signal, The algorithms used by the HIM module are the same used by Parini et al [7] and allow to extract from the EEG signal four different commands. HIM communicate via TCP/IP with the AEnima modules dedicated to the graphics interface. The visual stimulation for the SSVEP paradigms consists of four green LEDs placed one on each side of a 19'' LCD monitor. In this way each simulator can be associated with a 2-dimensional direction (LEFT, RIGHT, UP and DOWN) which correspond to the four BCI commands. BCI++ is based on a graphics engine [9] allowing to create an immersive graphic interface for the Raven test. Figure 3 shows the interface that for the test has been designed with two main features:

- be as similar as possible to the classic paper test;
- be used with only four commands.

Our computer implementation of the Raven test replicates the structure of the classic paper test. The possible solutions of the puzzle are presented in a matrix in which the subject can move with BCI commands. The selection method can be summarized as follows:

- The puzzle is displayed on the screen: the main picture is shown in the upper half of the screen while the six solutions are in the bottom half. The system is freezed for five seconds; the subject can't give any command. This time is left for identifying the correct answer.
- The software enables the LEFT and RIGHT commands of the BCI system. The subject can move through the possible solutions and selects the desired one.
- To select a solution the subject must stay on the same figure for more than three seconds. After this time a shadow border appears around the selected figure and the UP command is enabled.
- The user confirms his/her selection giving the UP command with BCI. He/she can still change the answer by moving the selection with LEFT/RIGHT commands.

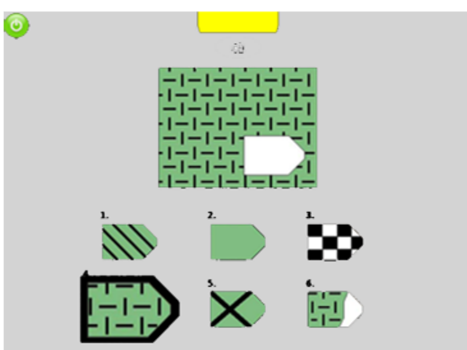


Fig. 3. BCI++ Graphics Interface for Raven Test

This selection method implements a double confirmation. We developed this policy in order to avoid incorrect answer due to a low bit-rate (Motor Imagery and SCP based BCI) which can affect test result. The software is able to record not only the answers but also the path the user made to reach the solution, the time to select the answer with BCI and the time to think about the solution.

2.3 The Participants

A group of 26 subjects (18 males, 8 females) aged from 10 to 41 years old participated to the study. The subjects are divided by pathology:

- 10 CPs (Cerebral Palsy) affected by paresis at different level;
- 1 affected by Multiple Sclerosis;
- 1 affected by Spinal Muscular Atrophy;
- 1 affected by dystonia
- 7 affected by Familial Spastic Para-paresis;
- 5 affected by Tetra-paresis;
- 2 affected by pathologies related to spinal cord injury and disease.

They had no previous BCI experiences. Every subjects with impaired vision performed the whole experimental session wearing the appropriate corrective lenses. All the measurement were carried out at the IRCCS E. Medea in Bosisio Parini (Lecco – Italy); written informed consent was obtained from the subjects and their families according to the Declaration of Helsinki. The research was approved by the Ethics Committee of IRCCS "E. Medea" Institute.

3 Results

26 pathological subjects between 10 and 41 years old take place to the study. Six subjects were excluded by the first exclusion criteria: one of the subject was excluded because affected by involuntary movements caused by dystrophy; two were excluded due to high electrode impedances which prevented the correct measurement of the signal; the last three were excluded because uncooperative.

Six subjects were excluded according to the second exclusion criteria: four of them didn't elicit SSVEP response while the other two have a pattern too small to be classified and used for the test (they are able to use less than four commands).

All the subjects who passed the first three criteria were able to complete the entire protocol. Table 1 reports the data for the 15 subjects who completed the test.

All the subjects completed the test in less than 15 minutes in accordance with which suggested by the test manual [11]. The bit-rate from 10 to 41 bit/sec shows how the BCI didn't affect test results. A bit-rate of 10 bit/sec is typical of P300 and Motor Imagery [1], this proves that our protocol can be used also with other BCI paradigms.

Table 1. Data for each subject who completed test; Bitrate is computed based on Walpow [10]

ID	Age	Sex	Bitrate	Raven	ZPoint	Test Time
AP	28	M	22,75	25	-1,78	370,77
EL	**39**	**M**	**10,50**	**23**	**-2,05**	392,46
IB	21	F	27,10	21	-2,26	371,09
PZ	**39**	**M**	**24,59**	**30**	**-0,30**	281,94
MA	8	M	39,52	30	0,20	198,54
GA	**14**	**M**	**41,62**	**31**	**0,50**	155,84
GG	14	M	24,02	31	0,50	303,43
VM	**12**	**M**	**36,48**	**30**	**0,20**	125,32
LT	15	F	30,32	35	1,69	175,92
CB	**22**	**F**	**18,19**	**31**	**0,07**	698,5
VV	20	F	22,89	30	-0,16	388,93
RS	**12**	**M**	**23,51**	**33**	**1,10**	230,11
AB	12	M	14,10	32	0,80	332,38
LF	**14**	**M**	**18,67**	**18**	**-3,38**	458,5
MM	12	F	11,55	35	1,69	448,96

4 Conclusion

Our study demonstrated that the here proposed protocol, already validated on healthy subject, can be used also on disable people for the administration of cognitive test. The BCI administration method, compared to the classical paper or computerized test, allows for the evaluation of new information which can be useful to study more complex aspects of cognitive abilities like selection strategy and spatial processing.

Our software was used with SSVEP based BCI but can be easily adapted to other BCI paradigms (eg. P300 and Motor Imagery) or other tests, and can be adopted, thanks to the use of already validated and widely used cognitive test, in clinical study were the administration of these kind of tests is difficult or, in some cases, impossible.

Acknowledgments. This work was partially supported by the Italian Institute of Technology (IIT). The authors thank doctors G. Livetti, L. Piccinini and engineers L. Maggi and S. Parini for their precious help and support, and all the subjects for their willingness and patience.

References

1. Wolpawa, J.R., Birbaumer, N., McFarlanda, D.J., Pfurtscheller, G., Vaughana, T.M.: Brain–computer interfaces for communication and control. Clinical Neurophysiology 113, 767–791 (2002)
2. Vidal, J.: Toward direct brain-computer communication. Annual Review of Biophysics and Bioengineering 2, 157–180 (1973)
3. Holzner, C., Guger, C., Grönegress, C., Edlinger, G., Slater, M.: Using a P300 Brain Computer Interface for Smart Home Control. In: World Congress on Medical Physics and Biomedical Engineering (2009)

4. Münßinger, J.I., Halder, S., Kleih, S.C., Furdea, A., Raco, V., Hösle, A., Kübler, A.: Brain Painting: First Evaluation of a New Brain-Computer Interface Application with ALS-Patients and Healthy Volunteers. Frontiers in Neuroscience 4, 182 (2010)
5. Brumberg, J.S., Nieto-Castanon, A., Kennedy, P.R., Guenther, F.H.: Brain-Computer Interfaces for Speech Communication. Speech Communication 52(4), 367–379 (2010)
6. Iversen, I.H., Ghanayim, N., Kübler, A., Neumann, N., Birbaumer, N., Kaiser, J.: A brain – computer interface tool to assess cognitive functions in completely paralyzed patients with amyotrophic lateral sclerosis. Clinical Neurophysiology 119, 2214–2223 (2008)
7. Parini, S., Maggi, L., Turconi, A.C., Andreoni, G.: A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication. Computational Intelligence and Neuroscience (2009)
8. Vialatte, F.B., Maurice, M., Dauwels, J., Cichocki, A.: Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Progress in Neurobiology 90(4), 418–438 (2010)
9. Perego, P., Maggi, L., Parini, S., Andreoni, G.: BCI++: a new framework for Brain Computer Interface application. In: International Conference on Computers and their Applications – Software Engineering and Data Engineering ISCA-SEDE (2009)
10. Kronegg, J., Voloshynovskiy, S., Pun, T.: Analysis of bit-rate definitions for Brain-Computer Interfaces. In: International Conference on Human-Computer Interaction (2005)
11. Rave, J., Court, J.H.: Manual for Raven's Progressive Matrices and Vocaboulary Scales. Assessment Harcourt (2003)

An Inductive Inference Model to Elicit Noncompensatory Judgment Strategies

Jing Yin[1] and Ling Rothrock[2]

[1] MicroStrategy Inc., 1850 Towers Crescent Plaza,
Vienna, VA 22182, USA
[2] The Pennsylvania State University, 310 Leonhard Building,
University Park, PA 16802, USA
jyin@microstrategy.com, lrothroc@psu.edu

Abstract. The proposed research developed a noncompensatory policy capturing technique to infer judgment rules (represented in disjunctive normal form) from available human data. The rule induction algorithm employs multiobjective Genetic Algorithm (GA) as its central search mechanism to enhance the induction and classification process. The quality of the induced rule set is measured by two criteria, *fidelity* (the degree to which the rule set reflects the judgment data they have been extracted from) and *compactness* (the simplicity of the rule set). An experimental study is conducted to demonstrate the effectiveness of the algorithm on a number of benchmark datasets.

1 Introduction

Von Neumann and Morgenstern [1] developed the classic Utility Theory which offers an idealized representation of decision outcomes. In their theory, it is assumed that people are rational decision makers who are capable of selecting the alternative with the maximum utility by using compensatory strategies. Though the classic Utility Theory dominates the literature for building normative decision models, critics argued that the Utility model had serious shortcomings [2]-[4]. Typically, when people make everyday decisions, they are not completely rational and consistent. They are not even aware of cue information that is readily available in the environment. Therefore, alternative models are needed to describe human decision making that requires less information processing.

Simon [3] proposed an alternative model based on the idea of satisficing, which suggests that people tend to select the first option that works for the situation rather than the optimal alternative. He later proposed the notion of bounded rationality [5] and stated that property of an agent that behaves in a manner that is nearly optimal with respect to its goals as its resources will allow. Noncompensatory decision models are developed to describe decision making behaviors that are consistent with the concept of bounded rationality. Noncompensatory strategies can be used to reduce the number of alternatives to be carefully evaluated and therefore improve the processing efficiency [6]-[9].

Regarding modeling noncompensatory strategies, although linear additive models can often approximate nonlinear strategies, they are sometimes inadequate to

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 414–422, 2011.
© Springer-Verlag Berlin Heidelberg 2011

represent rule-based behaviors, e.g., the XOR problem that is naturally arisen in real-world data [10] cannot be modeled by the weighting-and-summing formula [11] – it yields a set of zero regression coefficients for the predictor variables. As such, a need exists to develop a noncompensatory policy capturing model which can elicit heuristic rules from judgment data and does not rely on the linear compensatory assumption. The development of such systems is also of practical importance because it allows investigators to infer unobserved decision strategies from judgment profiles [12]. Rothrock and Kirlik [13] proposed that in a time-stressed, information-rich task environment, performers tend to develop heuristic rule-based strategies to cope with the increased cognitive workload. Therefore, rule-based modeling would be more appropriate than the tradition linear models in terms of describing the heuristic-based judgment behaviors. They thus presented a technique, called Genetics-Based Policy Capturing (GBPC), for inferring noncompensatory rule-based heuristics from judgment data, as an alternative to regression. The proposed research presented in this study serves as an advancement of the GBPC method and we intend to develop a technique that is more effective and efficient than GBPC.

2 The Inductive Inference Model

The proposed model uses inductive inference principles to generate rules in disjunctive normal form that are consistent with human decision behaviors. The current formulation of the problem builds on the existing work of GBPC [13]-[14] and extends it to accommodate discrete values with multiple categories.

2.1 General Problem Description

The general problem is that the decision-maker selects among *Q* alternatives $\{Y^1, Y^2, \dots, Y^Q\}$ regarding the states of the environmental criterion Y_e based on a set of environment cues $x_1, x_2, \dots, x_p$. We assume the i^{th} categorical cue, x_i can take categorical values from the domain set $\{X_i^1, X_i^2, \dots, X_i^{C_i}\}$, where C_i represents the number of categories for the i^{th} cue and C_i is finite. In addition, we specify that the rule-based model is represented in the disjunctive normal form which is expressed as the disjunction of *M* conjunctive rules:

To infer such rules from the judgment data, we assume the learning is based on *N* judgment instance profiles, which are shown in 2.2,

The exemplar matrix ***E***, summarizes the information of *N* judgment profiles. Each judgment profile is called an exemplar, which is represented by $e_{i\cdot}$. ***E*** is a two-dimensional matrix of size $N \times (p+1)$ in the form of:

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} & a_1 \\ e_{21} & e_{22} & \cdots & e_{2p} & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ e_{N1} & e_{N2} & \cdots & e_{Np} & a_N \end{bmatrix} \quad (1)$$

The element in the exemplar matrix e_{ij} for $i = 1,2,\ldots,N$, $j = 1,2,\ldots,p$ represents the value of the j^{th} cue in the i^{th} exemplar. The element a_i for $i = 1,2,\ldots,N$ represents the judgment value of the i^{th} exemplar.

Similarly, the rule set can be summarized via a two-dimensional matrix of size $M \times (p+1)$. We call it the rule set matrix, ***S***, where each conjunctive rule is represented by $s_{k\cdot}$.

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} & f_1 \\ s_{21} & s_{22} & \cdots & s_{2p} & f_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{Mp} & f_M \end{bmatrix} \quad (2)$$

The element of the rule set matrix s_{kj} for $k = 1,2,\ldots,M$, $j = 1,2,\ldots,p$ represents the constraint posed on the j^{th} cue in the k^{th} rule. s_{kj} can take any one value from the set. When $s_{kj} = 0$, the j^{th} cue can take any value from the set $\{X_j^1, X_j^2, \ldots, X_j^{e_j}\}$ for its condition to be met in the k^{th} rule. In addition, the element of the rule set matrix f_k for $k = 1,2,\ldots,M$ represents the decision value for the k^{th} rule.

2.2 Fitness of the Rule Set

The objective of the rule inference system is to elicit noncompensatory rules from judgment data. Therefore the quality of the rule set is determined based on its ability to represent noncompensatory decision making behaviors. The evaluation of rule sets is performed in two aspects, namely *fidelity* and *compactness*. First, *fidelity* stands for the degree to which the rule set reflects the judgment profiles from which they have been extracted. On the one hand, we want to generate accurate rule set to correctly classify the exemplar set. On the other hand, we want to minimize the classification error. Hence, *accuracy* and *error* are developed as two measures which correspond to *fidelity*. Second, *compactness* measures the simplicity of the rule sets. Rule set with fewer and simpler rules reflects a *satisticing* mode of decision making which is better suited to represent the noncompensatory decision strategies where less information is considered, integrated, and combined to make a judgment. The compactness of the rule set is measured by its *size* and *parsimony*. *Specificity* is introduced to the measure the compactness of a single rule.

All evaluation criteria are introduced in an attempt to improve the psychological plausibility of the induced rule set. We submit that the proposed approach does not aim to generate rule set to exactly mimic the decision making process but rather the goal is to generate a plausible hypothesis that is relevant to the use of noncompensatory decision strategies. In the next section we present the detailed mathematical representation of the fitness dimensions for the rule set.

2.3 Mathematical Representation of Fitness Dimensions

To compare the match between an exemplar $e_{i\cdot}$ with a rule $s_{k\cdot}$, we define a number of indictor variables. First, Table 1 lists all possible cases for the values of e_{ij} and

s_{kj}, and the match indictor α_{ikj} is defined as 1 if the value of the j^{th} cue in the exemplar is compatible with the value in the rule and 0 otherwise.

Table 1. Match between e_{ij} and s_{kj}

$s_{kj} = 0$	$e_{ij} = s_{kj}$	compatible	$\alpha_{ikj} = 1$
	$e_{ij} \neq s_{kj}$	compatible	$\alpha_{ikj} = 1$
$s_{kj} \neq 0$	$e_{ij} = s_{kj}$	compatible	$\alpha_{ikj} = 1$
	$e_{ij} \neq s_{kj}$	incompatible	$\alpha_{ikj} = 0$

Similarly for the decision value, Table 2 defines the match indictor β_{ik}. It is 1 if the decision values in the exemplar and the rule are the same and 0 otherwise.

Table 2. Match between a_i and j_k.

$a_i = j_k$	same	$\beta_{ik} = 1$
$a_i \neq j_k$	different	$\beta_{ik} = 0$

$$M_{ik} = [\alpha_{ik1} \quad \alpha_{ik2} \quad \cdots \quad \alpha_{ikp} \quad \beta_{ik}] \tag{3}$$

$$au_{ik} = \beta_{ik} \prod_{j=1}^{p} \alpha_{ikj} \tag{4}$$

Therefore, the matching vector of exemplar $e_{i\cdot}$ with respect to a rule $s_{k\cdot}$ is defined as:

Accuracy. The *accuracy* dimension is based on work in [15]. It measures the percentage of exemplars that are correctly classified by a rule set. The accuracy of using a single rule $s_{k\cdot}$ to classify an individual exemplar $e_{i\cdot}$ is defined based on the element of M_{ik}:

The accuracy score is 1 if and only if each element in the exemplar matches the corresponding element in the rule, and 0 otherwise. The accuracy of using the entire

$$au_i = \max_{k=1,2,\ldots,M} \{au_{ik}\} \tag{5}$$

rule set $\boldsymbol{S}$ to classify exemplar $e_{i\cdot}$ is the maximum accuracy of using any rule within rule set $\boldsymbol{S}$ to classify exemplar $e_{i\cdot}$:

$$au = \frac{\sum_{i=1}^{N} au_i}{N} \tag{6}$$

Then the accuracy score $\boldsymbol{au}$ $(0 \leq au \leq 1)$ of using rule set $\boldsymbol{S}$ to classify the exemplar set $\boldsymbol{E}$ is:

Error. *Error* dimension measures the percentage of exemplars that are wrong classified by a rule set. To determine the error rate, we first need to define the coverage of an exemplar $e_{i\cdot}$ matched by a rule $s_{k\cdot}$, which is based on the element of

$$c_{ik} = \prod_{j=1}^{p} \alpha_{ikj} \tag{7}$$

M_{ik}:

A rule is said to completely cover an exemplar if all conditions for the rule are true for the exemplar. For an individual exemplar $e_{i\cdot}$, its coverage with respect to the entire rule set ***S*** is defined as the maximum coverage by any rule within the rule set:

$$c_i = \max_{k=1,2,\ldots,M} \{c_{ik}\} \tag{8}$$

$$er_i = 1 \Leftrightarrow (c_i = 1) \wedge (au_i = 0) \tag{9}$$

An exemplar is said to be wrongly classified by a rule set if there exists rules that can completely cover the exemplar but no rule can be used to correctly classify it. As such, we define the error rate of using a rule set ***S*** to classify an exemplar $e_{i\cdot}$ as

To extend the error rate measure to the entire exemplar set, we define the error of using rule set ***S*** to classify exemplar set ***E*** as the percentage of exemplars that are

$$er = \frac{\sum_{i=1}^{N} er_i}{N} \tag{10}$$

wrongly classify by rule set ***S***:

Size. The size of a rule set is simply the number of rules it contains. For a rule set ***S***, its size is defined as

$$sz = M \tag{11}$$

Specificity. The specificity dimension was first suggested by Holland [16]. Rule sets with less match-all characters are classified as more specific [13]-[14]. For the k^{th}

$$sp_k = \frac{\sum_{j=1}^{p} y_{kj}}{p} \tag{12}$$

rule, if we have *r* cues specified that they can take any values, then the specificity for rule $s_{k\cdot}$ is 1 – *r*/*p*. Mathematically, by introducing another indictor variable y_{kj}, we can represent the specificity of $s_{k\cdot}$ as,

where y_{kj} is defined as:

$$y_{kj} = \begin{cases} 1, & s_{kj} \neq 0 \\ 0, & s_{kj} = 0 \end{cases} \tag{13}$$

The specificity ***sp*** for rule set ***S*** is the average specificity across the *M* rules. It is defined as:

$$sp = \frac{\sum_{k=1}^{M} sp_k}{M} \tag{14}$$

From the formulation we know $0 \le sp \le 1$. Although in general we want to minimize the specificity of the rule set to generate simpler and more general rules, we keep in mind that an over generalized strategy[1] should be avoided at all times since it represents a complete blind and random judgment behavior which provides zero information to our subject of interest. Therefore we would not want to incur such types of rules in the candidate rule sets and they should be of minimum use.

$$ps_k = \frac{\sum_{i=1}^{N} au_{ik}}{N} \tag{15}$$

Parsimony. The parsimony dimension reflects the goodness of a rule set in terms of the necessity of each rule [13]. More specifically, it measures the strength of each rule in the rule set in terms of its utility to classify the exemplars. The utility of using s_k- to classify the entire set of exemplars ***E*** is defined as:

The parsimony level ***ps*** $(0 \le ps \le 1)$ of a rule set ***S*** is the maximum utility scores found across the *M* rules:

$$ps = \max_{k=1,2,\ldots,M} \{ps_k\} \tag{16}$$

2.4 Multi-Objective Optimization GA Algorithm

The rule extraction problem can be then formulated as a five-objective optimization problem. The first two objectives guarantee that we solve the problem accurately, while the last three introduce pressure towards compact solutions. Given that we have finite numbers of rule sets ***S*** whose length varies from 1 to *M*[2], we want to find the set of rule sets that are *Pareto* optimal with respect to the five dimensions (***au***, ***er***, ***sz***, ***sp***, ***ps***).

Based on the review of the existing techniques, we found that a method that can explicitly deal with the criteria we have defined in Section 2.3 has not been fully developed yet. This brings the necessity of developing an appropriate rule induction algorithm that can be used to model noncompensatory decisions. In this study we explore the use of multiobjective GA to construct a system which can learn noncompensatory rules from human decision data. The system can also be used to solve a wide variety of machine learning classification problems with various configurations.

[1] An over generalized strategy is exhibited by the inclusion of a rule in the form of IF (CUE 1 = ANYTHING) AND (CUE 2 = ANYTHING) AND … AND (CUE P = ANYTHING) THEN (SELECT ALTERNATIVE J). This translates to the vector representation of having zeros values for all cues.

[2] *M* is the maximum number of rules that is allowed to be included in a rule set ($M \le N$). It could be a psychologically relevant variable that defines the capacity of a human in terms of the number of rules he/she can manipulate for his/her judgment strategy.

The empirical testing results are provided in the next section to illustrate the system performance in relative to existing methods.

Our proposed method is built on the basis of GBPC. Empirical testing on GPBC shows that the system performance is very sensitive to the initial populations and the system can produce drastically different results across runs. In addition, the solution quality for large size problems is not on par with the current state-of-the-art rule induction algorithms. In order to improve the effectiveness of the algorithm, we propose a hierarchical system where GA works on two levels as in [17]. The technique essentially performs an initial partition on the solution search space and divides it into many sub regions. Then knowledge is learned on each of these sub regions. We expect that the newly developed technique can greatly improve the performance of the algorithm and produce consistent results over times.

3 Experimental Study

The purpose of the experimental study is to test the effectiveness of the proposed rule induction algorithm. In this experiment, we will compare the results of our algorithm with the results produced from a number of well known learning algorithms including ID3, C4.5, k-NN, IB1, and naïve Bayes. We abbreviate our algorithm as RI for the rest of this paper. The algorithm itself is implemented in Matlab® version 7.0 Release 14.

Two measures used in this study to rank the performance of the learning algorithms are accuracy and compactness. The higher the accuracy is, the better the model generalizes from the data. Compactness measures the simplicity of the model where in our case refers to the size of the rule set. An ideal rule set will have minimum number of rules and correctly classify all instances in the test set.

In our study, the testing results on nine UCI datasets including balance scale, car evaluation, congressional voting records, hayes-roth, lenses, monk-I, monk-II, SPECT heart, and tic-tac-toe endgame are reported. The UCI (University of California Irvine) machine learning repository is a collection of databases used by practitioners and researchers from all over the world to conduct empirical analysis of machine learning algorithms [18]. Results reported regarding the performance of the benchmark algorithms on the UCI datasets are those that have been found in the existing machine learning literature. These results were generated using the same validation technique as we used for all the datasets and therefore the comparison is valid. However, only partial results are available for some of the datasets.

Overall, our results show that RI generated simple yet accurate rules as compared with other benchmark algorithms on many of the tested datasets if not most. It suited particularly well on datasets generated using disjunctive functions. However, as literature suggested that no algorithm outperforms others in all settings and domains [19], RI did produce inferior results on a small portion of the tested datasets. In conclusion of the experimental study, we submit that RI is an effective rule induction algorithm which performs comparably well with many state-of-the-art machine learning algorithms.

4 Conclusions

In this study, we used zero order predicate rules to represent the learned knowledge and avoided the combinatorial complexity in computation. In future studies, higher order representation could be introduced to the model and we believe this added flexibility in representation could improve the system performance in terms of its robustness on a broader range of domains.

References

1. Neumann, J.v., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1947)
2. Mosteller, F., Nogee, P.: An experimental measurement of utility. Journal of Political Economy 59 (1951)
3. Simon, H.A.: A behavior model of rational choice. Quarterly Journal of Economics 69, 99–118 (1955)
4. Luce, R.D.: Semi-order and a theory of utility discrimination. Econometrica 24, 178–191 (1956)
5. Simon, H.A.: Models of Man. Wiley, New York (1957)
6. Payne, J.W.: Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance 16, 366–387 (1976)
7. Beach, L.R.: Measures of linear and nonlinear models of decision behavior: Process tracing versus policy capturing. Organizational Behavior and Human Performance 31, 331–351 (1993)
8. Payne, J.W., Bettman, J.R., Johnson, E.J.: The Adaptive Decision Maker. Cambridge University Press, New York (1993)
9. Payne, J.W., Bettman, J.R., Luce, M.F.: Behavioral decision research: An overview. In: Birnbaum, M.H. (ed.) Measurement, Judgment, and Decision Making, pp. 303–359. Academic Press, New York (1998)
10. Page, D., Ray, S.: Skewing: An efficient alternative to look ahead for decision tree induction. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 601–607 (2003)
11. Yin, J., Rothrock, L.: A rule-based lens model. International Journal of Industrial Ergonomics 36, 499–509 (2006)
12. Elrod, T., Johnson, R.D., White, J.: A new integrated model of noncompensatory and compensatory decision strategies. Organizational Behavior and Human Decision Processes 95, 1–19 (2004)
13. Rothrock, L., Kirlik, A.: Inferring rule-based strategies in dynamic judgment tasks: Toward a noncompensatory formulation of the Lens Model. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 33, 58–72 (2003)
14. Rothrock, L., Ventura, J., Park, S.: An optimization methodology to investigate operator impact on quality of service. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando (2003)
15. DeJong, K.A., Spears, W.M., Gordon, D.F.: Using genetic algorithms for concept learning. Machine Learning 13, 161–188 (1993)

16. Holland, J.H., Holyoak, K.F., Nisbett, R.E., Thagard, P.R.: Induction. The MIT Press, Cambridge (1986)
17. Markowsha-Kaczmar, U., Mularczyk, K.: GA-based Pareto optimization for rule extraction from neural networks. Studies in Computational Intelligence 16, 313–338 (2006)
18. Blake, C., Keogh, E., Merz, C.J.: UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, CA (1998), http://www.ics.uci.edu/~mlean/MLRepository.htm
19. Schaffer, C.: A conservation law for generalization performance. In: Machine Learning: Proceedings of the Eleventh International Conference, pp. 259–265. Morgan Kaufmann, San Francisco (1994)

A User-Friendly Tool for Detecting the Stress Level in a Person's Daily Life

Irene Zaragozá[1,4], Beatriz Rey[1], Cristina Botella[2,4], Rosa Baños[3,4], Inés Moragrega[2], Diana Castilla[2], and Mariano Alcañiz[1,4]

[1] Instituto InterUniversitario de Investigacióm en Bioingeniería y Tecnología Orientada al ser Humano, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain
[2] Universitat Jaume I, Castellón, Spain
{botella,moragrega,castilla}@psb.uji.es
[3] Universidad de Valencia, Spain
[4] CIBER Fisiopatologia Obesidad y nutrición (CB06/03), Instittuto Carlos III, Spain
{izaragoza,brey,malcaniz}@labhuman.i3bh.es, banos@uv.es

Abstract. Mental health care represents over a third of the cost of health care to all EU nations and, in USA, it is estimated to be around the 2.5% of the gross national product. Depression and Stress related disorders are the most common mental illnesses. The European project OPTIMI will develop tools to make predictions through the early identification on the onset of the disease. In this paper, we present a user-friendly application developed in the OPTIMI project to detect the stress level in a person's daily life. The results of a first usability study of this application are also presented.

Keywords: mental health, usability, depression, stress detection, prevention.

1 Introduction

Mental health problems account for more than a third of the cost of health care systems of all European Union countries [1]. They also have additional costs in economic terms due to loss of productivity. Depression and stress related disorders are the most common mental illnesses and the prevention of depression and suicide is one of the five key points in the European Pact on Mental Health and well-being.

According to European Study of the Epidemiology of Mental Disorders (ESEMeD) project, major depression and specific phobia were the most common single mental disorders in the EU.

While depression affects one in six people (in the UK) at some point in their life and is most prevalent among people aged 25-44, it is increasingly affecting adolescents. In Europe, 8% of all girls and 2% of all boys show symptoms of severe depression [1]. In the UK, among the 5-10 age group, 10% of boys and 6% of girls are affected; and among the 11-15 age group, 13% of boys and 10% of girls.

The worrying figures confirm that we need to develop better treatments to help those affected, and more importantly, given the near exponential explosion in the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 423–431, 2011.
© Springer-Verlag Berlin Heidelberg 2011

number of sufferers (young, adolescent, adult, aged and repeat episodes), we need to develop better means for identification of subjects at risk and for prevention, in particular as adolescents move on into the more stressful environments of the workplace and bringing up a family.

Currently, the main treatments for mental illnesses are pharmaceutical drugs based on evidence and cognitive-behavioural therapy (CBT) [2-3]. However, little is being done to develop effective systems for preventing the onset of these diseases. Many of us are lucky not to be subject to daily stress conditions that ultimately lead to physical and personality changes. Furthermore, some people are lucky enough to cope with high pressures. However, many find themselves in high-risk situations where, despite their efforts, they decompensate and develop a depressive disorder.

As previously commented, Depression and General Anxiety Disorders (GAD) are treatable illnesses with CBT as the treatment of choice. However, due to various reasons, the provision of mental health care is generally less than adequate in terms of accessibility and quality. CBT comprises a set of therapist and patient processes whose format allows for the whole treatment process to be computerized and personalized, Computerized CBT (CCBT). For example, in the UK, the Netherlands, Sweden, Italy, Spain, as well as many centres in the USA and Canada there are now emerging several CCBT treatments that include the wide use of Internet and CD based solutions [4-7]. The mentioned CCBT treatments provide solutions to disorders such as: fear of public speaking [4], obsessive-compulsive disorder [5], panic disorder [6] or anxiety and depression [7]- These computerized treatments are being actively tested and, in certain cases, licensed for wide application to meet the growing need for treatment.

So CCBT will provide a good alternative, for traditional CBT that is restricted due to a combination of the number of available therapists, with adequate quality training, as well as economics and physical accessibility.

1.1 The Optimi Project

However while the extensive computerization of CBT at all levels will indeed enhance the CBT treatments and provide a powerful means for effective deployment in general mental health care provision, there is something lacking in the computerization that is sadly missing. What is missing are the tools that reliably predict and diagnose the onset of the illness, the development of the illness and the effects of treatment on the illness.

In order to identify the onset of mental illness, a tool for detecting the occurrence of stress in the daily life of an individual has been developed. This tool is part of a European project called OPTIMI (project number 248544) to determine the continuous effect of stress on the individual by studying patterns of behaviour over a longer period in order to predict whether the user is at risk of depression and, if so, it will recommend him a set of activities to reduce his stress level.

OPTIMI has set itself two goals: first, the development of new tools to monitor coping behaviour in individuals exposed to high levels of stress; second, the development of online interventions to improve this behaviour and reduce the incidence of depression. In this work, we will focus on the first goal.

In order to detect the stress level of a person, the following measures are considered: ECG [8], physical activity [9], voice analysis [10], EEG [11] and self- report diary [12].

Power spectrum analysis of heart beat for quantification of heart rate variability (HRV) is a non invasive method for identifying the activity of the autonomic nervous system (ANS) and its control functions. The HR/HRV technique uses these variations to assess physiological responses to various real world conditions in normal and diseased subjects.

OPTIMI uses actigraphy (physical activity) for two different purposes: (i) to analyze ambulatory activity, exercise and resting, (ii) to estimate sleep quality. Actigraphic measurements make possible to estimate to what extent subjects' subjectively experienced fatigue corresponds to objectively measured physical activity, to what extent it is the result of chronic stress and over activation of the sympathetic system, and to what extent it is a psychosomatic symptom.

The study of speech patterns of users is used by the application in order to establish their state of affect. The evolution of human voice enables speakers to non-verbally communicate information on their affective state, including states of mind (sadness, happiness, fear, anger, aggression, etc.), likely to be associated with mental and physical manifestations.

On the other hand, OPTIMI uses power spectrum analysis, time-frequency analysis, correlation dimension (D2) and Approximate Entropy (ApEn) analysis of scalp EEG to identify variations in brain functioning and emotional responses in subjects belonging to high risk groups.

Finally, self reporting strategies provide the best possible psychological information to the OPTIMI system related to the possible risk of developing a major depressive disorder; that is, the users' own self report of their experienced stress and how they are coping with it.

All these measurements have to be taken in a daily basis, as long as the variability over time and dynamic patterns of those variables will be fundamental to detect changes that are associated with increments in stress levels.

In this work, we will focus on the description of the user-friendly application that has been developed to allow the registration of the different signals and user data. The results of a preliminary usability study of this application will be presented and discussed.

2 Application Developed

From the user perspective, it is necessary that a simple and intuitive tool that does not interfere in the user's daily life is used to monitor all these data. In psychological research, there is a growing trend to use computer programs that can prompt the subjects on a periodic base to rate and give answers about different parameters [13].

In the OPTIMI project, the tool developed for detecting the stress level in a person's daily life presents a user-friendly interface that allows the user to obtain the data collected by the activity and ECG sensors, to perform specific measurements such as voice analysis and EEG processing and to complete a self- report diary.

The ECG and activity sensors are worn by the user during the whole day, and the application is used to download the registered data. The computer automatically interfaces any nearby sensor wirelessly and downloads any waiting processed information. The computer uses normal secure internet connections to send the data to the OPTIMI server. A complete description of these sensors can be found elsewhere [14].

The application has been developed in C++ and it is a local application that must be installed in the PC before it is handed over the final user. It is a multilanguage application in which the language will be detected automatically based on the language of the operating system.

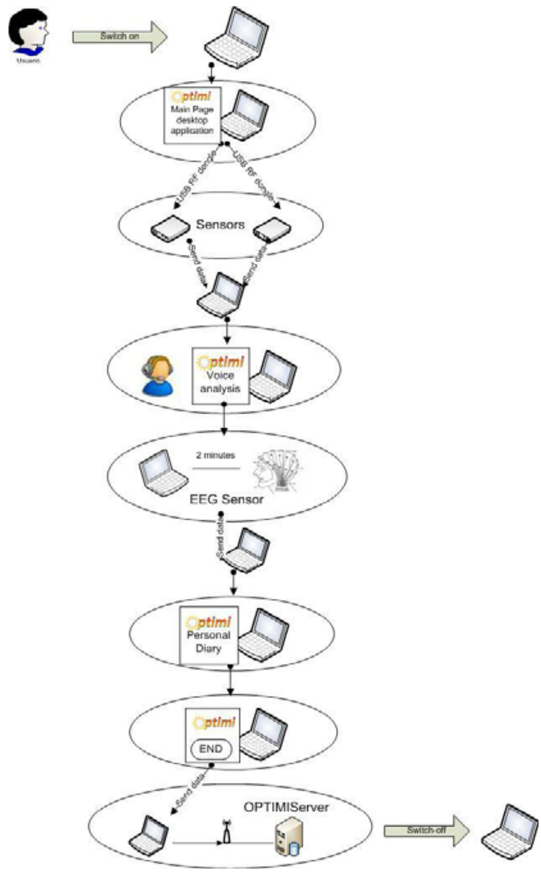


Fig. 1. Daily routine of the application

2.1 Usage Model

The activity and ECG sensors are worn more or less continuously (24/7). Of course they may be removed for comfort and will need to be removed when the batteries need recharging. They can then be swapped for an identical spare unit.

Once a day the user is requested to go to the PC and switch it on. At this point the PC should automatically begin the application and a process of downloading data from the ECG and Activity sensors as well as interacting with the user to obtain an EEG reading, a Speech reading and finally a Self report diary reading.

Once all data has been provided the user may walk away. The PC then performs some final computations, and contacts the OPTIMI server via the Internet. Finally it will shutdown gracefully and switch off automatically. Fig. 1 shows the daily process diagrammatically.

2.2 Workflow

The application has a wizard look and feel that guides the user through the different steps in order to minimize possible errors. The communication with the different sensors is done through different dlls. Following the different steps the application are described:

1. The user switches the computer on. The Optimi application is launched automatically and the "Welcome screen" appears. The application blocks the computer and hides the windows toolbar, not allowing the user to close it.
2. The user introduces his pin number. In order to increase the security and to avoid possible communication problems the login process is done locally. The PC contains an encrypted configuration file that contains the UserId and the PIN number.
3. At this point the Home PC, via a USB Radio frequency dongle, searches for ECG and Activity sensors. If no sensor is detected, the application asks the user to stay with the sensors near the PC. Once the sensors are connected to the PC the application sends a message to the ECG and Activity sensors using a proprietary 2.4GhZ RF data stream to download their data. The data downloaded from each sensor is stored appropriately in a text file.
4. The application starts the voice analysis process indicating the user how to connect the microphones needed for this process. After this, the application asks the user to talk into the microphones following guided procedures. The different steps of the voice procedure are: count from 1 to 40, read a text passage that lasts 1 or 2 minutes and then count form 1 to 40 again. This process generates different wav files that are sent to the voice analyzer module to obtain a depression score of the user.
5. The application starts the EEG analysis module showing the user how to put the EEG sensor. After this, the application asks the user to breathe deeply and records the EEG signal for 85 seconds. This signal is downloaded from the EEG sensor via the USB RF dongle and sent to the EEG analysis module to obtain a stress score of the user.
6. At this point, the user has to fill in a self-dairy questionnaire indicating a value for the following variables: coping capability, daily mood, stressful events, general activity and general health. Moreover, the user gives information about the activity that has done during the day.
7. At this step, the application encapsulates all the information obtained in the previous points in a single package (zip file) and sends it to the server to its proper storage and analysis. After that the application shows a message to the user thanking his participation and the PC switches off automatically.

The application combines all the data obtained in the steps previously described to detect the stress level of the person in his daily life. It is an easy to use application that the user can apply every day to monitor his stress level.

Fig. 2. Some screenshots of the application

3 Usability Study

The usability and acceptability of the application is a key factor in the OPTIMI project. The participants in the project trials will use the application alone, at home without the help of the therapists, so it is very important they feel comfortable with it. For this reason, the application should be self-explaining and easy to use.

To test the acceptability and usability of the application a usability study has been conducted. The main objective of the study is to determine if the users are able to do the different tasks the application proposes (download data from sensors, record voice and fill in a self-diary questionnaire) and if they feel comfortable with the application. It was not an objective of this study to evaluate if the application was useful to evaluate stress levels in a person's daily life. That will be analyzed in future studies inside the OPTIMI project.

3.1 Participants

In the usability study participated 10 users (3 women and 7 men), aged between 25 and 69 years. Since the final application will be used by users of a wide age range, the study conducted intended to obtain sample of different ages.

Before starting the study the users were asked about how they felt when working with computers and new technologies, using the visual analogue scale (VAS) from 1 to 7 represented in Fig. 3. As shown in Table 1. most of the users felt comfortable working with computers.

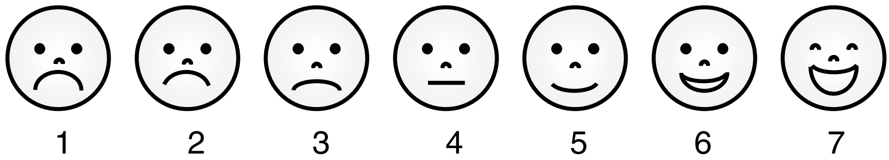


Fig. 3. Visual Analogue Scale used in the usability study

Table 1. Results of the pre usability study question

Visual analog scale	Percentage of answers
1	0%
2	0%
3	10%
4	20%
5	10%
6	30%
7	30%

3.2 Experimental Procedure

Following, the procedure used in the usability study of the application is described. First, the user is explained how the application works step by step. Secondly, the user is left alone with the computer to use the application. Finally, the user is asked some questions about his experience with the application and its usability.

To measure the usability of the system, the SUS (System Usability Scale) questionnaire is used [15]. The SUS covers the following aspects:

- Effectiveness (the ability of users to complete tasks using the system, and the quality of the output of those tasks)
- Efficiency (the level of resource consumed in performing tasks)
- Satisfaction (users' subjective reactions to using the system).

3.3 Results

As previously commented, the users complete the SUS questionnaire after having used the application. SUS is a simple, ten-item scale giving a global view of subjective assessments of usability. The principal value of the SUS is that it provides a single number (from 0 to 100) representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. Table 2. shows the results obtained in SUS questionnaire for each participant, the mean of all the scores and the standard deviation.

Taking into account the adjective rating scale presented in [16] and the mean obtained at the SUS questionnaire (73.5), our application could be considered as "good" in terms of usability.

Table 2. SUS scores

Participant	Sus score
User1	90
User2	77.5
User3	75
User4	70
User5	70
User6	77.5
User7	80
User8	52.5
User9	72.5
User10	70
Mean	**73.5**
Standard Deviation	**9.59**

Apart from the SUS questionnaire, the participants were asked about how they felt using the system with the same VAS scale used in the pre-study question. As shown in Table 3., users felt neutral or comfortable using the system.

Table 3. Results of the post usability study question

Visual analog scale	Percentage of answers
1	0%
2	0%
3	0%
4	40%
5	40%
6	10%
7	10%

4 Conclusions

The European project "Optimi" aims to develop a set of tools to prevent an individual from falling into depression. For this reason a friendly application that can detect the stress level of a user in his daily life has been developed. This application collects and combines different types of measures: ECG, physical activity, voice analysis, self-registration and EEG to obtain the individual's stress level.

An initial usability study of the application has been conducted and has been presented in this work. The results of this study show that the application has a friendly, easy to use interface and that users feel comfortable using the application. As has already been commented, the application will be used daily during the OPTIMI project during extended periods of time in order to evaluate stress levels in a person's daily life. Results of the present study seem to support the hypothesis that the use of the application during longer periods would not interfere in the user's routines. Then, it could be a contribution to the growing trend of applying computer programs to collect data about psychological aspects on a periodic base.

References

1. World Health Organization, Mental Health in the European Region. WHO, Geneva (2003)
2. Woody, S.R., Sanderson, W.C.: Manuals for Empirically Supported Treatments: 1998 Update. The Clinical Psychologist 51, 17–21 (1998)
3. Nathan, P.E., Gorman, J.M.: A guide to treatments that work, 2^{a}nd edn. Oxford University Press, Oxford (2002)
4. Botella, C., Quero, S., Banos, R., Garcia-Palacios, A., Breton-Lopez, J., Alcaniz, M., et al.: TelePsychology and Self-Help: the treatment of phobias using the Internet. CyberPsychology & Behavior 11(6), 659–664 (2008)
5. Greist, J., Marks, I., Baer, L., Kobak, K., Wenzel, K., Hirsch, M., et al.: Behavior therapy for obsessive-compulsive disorder guided by a computer or by a clinician compared with relaxation as a control. Journal of Clinical Psychiatry 63(2), 138–145 (2002)
6. Marks, I., Kenwright, M., McDonough, M., Whittaker, M., Mataix-Cols, D.: Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. Psychological Medicine 34(01), 917 (2004)
7. Proudfoot, J., Ryden, C., Everitt, B., Shapiro, D., Goldberg, D., Mann, A., et al.: Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. The British Journal of Psychiatry 185(1), 46 (2004)
8. Huang, M.H., Ebey, J., Wolf, S.: Responses of the QT Interval of the Electrocardiogram during Emotional Stress. Psychosomatic Medicine 51, 419–427 (1989)
9. Taliaferro, L.A., Rienzo, B.A., Pigg, R.M., Miller, M.D., Dodd, V.J.: Associations between physical activity and reduced rates of hopelessness, depression, and suicidal behavior among college students. Journal of American College Health 57(4), 427–436 (2009)
10. Russell, J.A., Bachorowski, J.A., Fernández-Dols, J.M.: Facial and vocal expression of emotion. Annual Review of Psychology 54, 329–349 (2003)
11. Hayashi. T, Mizuno-Matsumoto, Y., Okamoto, E., Ishii, R., Ukai, S., Shinosaki, K.: Anterior brain activities related to emotional stress. Presented in World Automation Congress, San Antonio, USA (September 2008)
12. Piasecki, T.M., Hufford, M.R., Solhan, M., Trull, T.J.: Assessing Clients in their natural environments with electronic diaries: Rationale, Benefits, Limitations, and Barriers. Psychological Assessment 19, 25–43 (2007)
13. Stiglmayr, C.E., Grathwol, T., Linehan, M.M., Ihorst, G., Fahrenberg, J., Bohus, M.: Aversive tension in patients with borderline personality disorder: a computer-based controlled field study. Acta Psychiatr Scand 111, 372–379 (2005)
14. Majoe, D., Kaegi-Trachsel, T., Gutknecht, J.: A Smart Wearable Sensor Assisting in Mental Health. Presented in 12 ACM International Conference on Ubiquitous Computing, Copenhagen, Denmark (September 2010)
15. Brooke, J.: SUS- A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor and Francis, London (1996)
16. Bangor, A., Kortum, P.T., Miller, J.T.: An Empirical Evaluation of the System Usability Scale. International Journla of Human Computer Interaction 24(6), 574–594 (2008)

"How Do I Line Up?": Reducing Mental Transformations to Improve Performance

Guy W. Zimmerman[1], Dale Klopfer[2], G. Michael Poor[1], Julie Barnes[1], Laura Leventhal[1], and Samuel D. Jaffee[2]

[1] Computer Science Department
[2] Psychology Department, Bowling Green State University,
Bowling Green, OH USA 43403
{gzimmer,klopfer,gmp jbarnes,leventha,jaffees}@bgsu.edu

Abstract. Mobile devices and visual-spatial presentations of information are pervasive, especially for tasks in which the mobile device can be moved to close proximity of the task. This mobility allows the user to offload mental workload by allowing physical transformations of the device. In this study, we compared a fixed mobile device, a non-fixed mobile device, and a fixed desktop display to determine the effects imposed by the mental workload of transforming the frames of reference into alignment. Our results indicate that allowing the user to manipulate the device's position can influence performance by reducing the need for mental transformations.

Keywords: Mobile Device, Usability, 3D Models, Frames of Reference.

1 Introduction

Mobile devices allow for the delivery of information for many applications in a diverse set of contexts. The very feature that defines mobile devices, that is their *mobility,* would seem to have great potential for modifying the mental demands required for a variety of tasks. Consider a person trying to watch a video while lying down. If they are watching the video on a standard television screen, they would probably prop their head up to get a better view of the video. However, if they were watching this video on a tablet computer, or some other mobile device, they could keep their head in a lying down position, while rotating the screen of the device to an orientation where the video could easily be viewed. Now suppose that a person is watching an interactive presentation with 3D objects on a mobile device, for example, a 'rotatable' origami whale. Once again, the user has the capability of changing the view of the scene by changing the position of the device or by changing their own position. However, by virtue of the interactivity of the presentation, they may also change their view of the scene by reorienting the whale within its virtual space (Fig. 1). Any of these changes might occur in an effort of the user to realign the presentation with the device, the device with the user and so on. If the device, the presentation, or the person are fixed, as in the case of the television, the user is forced to *mentally* realign salient elements of the user experience. It is this mental workload, imposed by the process of mental realignment, which would seem to be potentially reducible in a mobile device.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 432–440, 2011.
© Springer-Verlag Berlin Heidelberg 2011

In this paper we present results of a study of mobile device usability in which we vary the opportunities for the task presentation to impose mental workload via demands for mental or physical alignment. We first review the relevant literature, present our methodology and results.

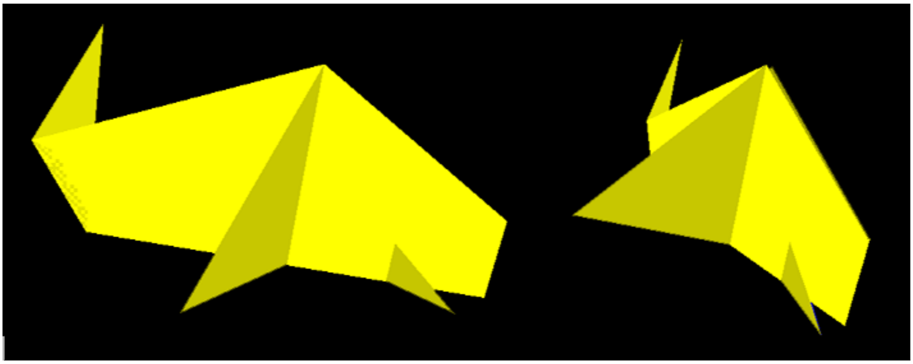


Fig. 1. Two views of a (rotated) origami whale within its virtual space

2 Background

When a person mentally realigns themselves and objects in their world, a major element of the resultant mental workload in is the management of frames of reference (FoRs). When a person encounters an object in the world, they must mentally manage various FoRs, including FoRs induced by: the object, the person, and gravity. Mental "management" means the mental workload of transforming the FoRs into alignment with the viewer [12, 13, 15]. When viewing an object in virtual space (i.e. in a computerized presentation), the relevant FoRs are those induced by: the viewer, the computer display, and the object in virtual space [c.f. 11, 16]. In our study, using an interactive presentation, we examine the effects imposed by the management of these FoRs with one additional factor: because our task was a multi-stage construction of an actual object, an additional FoR (of the constructed object) to manage was introduced. When a person and all elements in their environment are fixed (e.g. a person with their head restrained viewing objects on a fixed desktop computer screen), the person must mentally realign the relevant frames of reference for themselves. When any element of the person or the environment are not fixed (e.g. a person with free movement of their head and hands viewing objects on a non-fixed mobile device), the person has the opportunity to offload the mental work of realignment by physically moving themselves or salient environmental items.

In previous studies, we explored the effectiveness of 3D graphics as a component of a system to deliver instructions for a *construction task* [cf. 1, 16]. Construction tasks are ubiquitous and include activities like assembling a child's bike; the traditional form of these instructions has been written text accompanied with 2D illustrations. Typically, the task is divided into a number of stages with multiple steps per stage. Such instructions can be notoriously difficult to follow for a variety of reasons: the number of steps per stage can be "overwhelming" or the user can find it difficult to discern the actual 3D spatial relationships among the parts from the 2D illustration. Creating a true 3D representation (i.e. where the user can view the scene from any desired vantage point) of the construction process offers the potential to address these and other limitations of the traditional paper presentation style. Such presentations offer opportunities to study the role of FoRs because the richness of the applications induce the simultaneous existence of multiple FoRs.

In our prior studies we have examined usability issues of computerized presentation of instructions for folding origami objects across a variety of contexts. In this, or any origami task, the participant is required to construct a three-dimensional object through a series of folds, unfolds, and creases (much like the construction of a paper airplane). While no single task can be expected to singularly exemplify the wide range of real-life construction tasks, origami folding does possess many representative characteristics: the task is non-trivial, requires multiple manipulation steps, and results in a 3D artifact. Furthermore, its lack of confounding concerns (e.g. use of tools) makes it an almost ideal task for focusing on our specific research questions. Finally, from an experimental design view it is also nearly ideal: the task is familiar to participants (extensive training is not required), very low cost, very low risk, and allows for investigation of construction tasks of incremental complexity.

Our prior work included studying the relative effectiveness of various presentation formats: paper, computerized presentation of text with 2D illustrations, and computerized presentation with interactive 3D models; the latter on a variety of platforms, including mobile devices [1, 10, 16, 17]. Qualitative analysis of participant behavior in our mobile device experiments led us to consider the implications of the management of FoRs for this problem domain and to the particular study presented here. By nature, a desktop display is fixed where a mobile device is mobile and the user may alter its position/orientation. As a result, for applications involving 3D graphics, interactive and static, the mobility of the device potentially changes the way that the user visualizes the elements of the presentation and their relationships. Specifically, the mobility of the device potentially allows the user to see the presentation in different orientations or to see hidden content in ways different from a non-mobile device. Thus, at a very minimum, it is not clear to what extent (if any) usability issues and solutions for desktop applications map onto corresponding issues and solutions for mobile devices.

2.1 Evidence of the Importance of Presentation Orientation

Three very different studies point to the importance of orientation in a presentation of interactive visual information. [11] found that, for map-based navigation assistance, physical rotation is the most effective form of track-up alignment on handheld mobile devices [11] found users' difficulty recognizing a map when automatically rotated, especially when the users were not looking at the map during the time of rotation.

[14] describe a comparative study of the effectiveness of four different presentations of instructions for an assembly task: printed manual, monitor-display, see-through head-mounted display, and spatially registered augmented reality (AR). Measurements were task performance (time and accuracy) and perceived mental workload. The task consisted of 56 procedural steps building an object with Duplo blocks. Participants in the spatially registered AR treatment made significantly fewer

assembly errors. The authors concluded that the improvement in the AR condition was due to reduced demand for attention switching. Because the spatially registered AR appears directly on the object, it was also thought that the participants made fewer mental transformations between the instructions and the object.

[4] report on a design tool to build three-dimensional, interactive and movable polyhedrons. In evaluating this tool, they found that users had a preferred orientation for the designed polyhedrons. When the figures were moved from the preferred orientation, participants found them to be more difficult to sketch (reproduce by hand). Some reported elements of the preferred orientation include: 1) preference for vertical as opposed to horizontal edges (preference for either types of edges as compared to diagonal edges), 2) bilateral symmetry, and 3) stability as indicated by the polyhedron resting on a face as opposed to resting on a vertex.

3 Method and Procedure

According to prior research [2, 6, 13], mental workload increases and performance decreases with mental management of increasing numbers of FoRs. Construction tasks, by their nature, often require users to manage and align several FoRs. Deploying instructions in a 3D presentation offers users a way to physically align FoRs within the presentation but when the 3D presentation is shown on a fixed display, that very property of the paper instructions, physical realignment of the paper, is lost. However, both opportunities for physically realigning the FoRs of the tasks can be regained with a 3D presentation on some sort of mobile device. The purpose of the current study is to test the hypothesis that a mobile presentation of instructions for a construction task leads to better performance than is achieved when the presentation is on a fixed device, by virtue of offering opportunities for physical as opposed to mental realignment of FoRs.

In the current study, participants could reduce mental workload by aligning one or more FoRs; e.g. participants could align the constructed object with the presentation on the screen, the constructed object with the screen itself, or in some cases, the screen with the constructed object. In all conditions of the present study, participants could mentally or virtually (via an interactive presentation) realign any of the above FoRs. However, participants in the non-fixed mobile device condition could also physically realign the device itself to reduce mental workload. To clarify, in all conditions, participants could rotate the presentation of the instructions in virtual space to align the FoR of the virtual object to reduce mental workload. Also, all participants could, of course, realign the constructed origami whale itself to any orientation to possibly reduce mental workload. If participants performed no realignments in the virtual or physical world, it was assumed they performed realignments mentally, thus increasing mental workload [as evidenced by 2, 6, 13].

We compared two conditions with varying demands for management of FoRs, a fixed mobile device in a cradle and a non-fixed mobile device. FoRs of the non-fixed device could be aligned physically or mentally. We added a third condition for external validity: a fixed desktop display. In all conditions, the presentation contained virtual interactive instructions for constructing an origami whale and a paper packet of line drawings and text.

The current study was actually part of a larger study [10] that considered additional issues not germane to the current discussion. We examined three dependent performance variables: the proportions of correct folds (PCF), recreased folds (PRF) (correct but redone folds) and incorrect folds (PIF); if a participant performed a step incorrectly, then corrected the error later, these were two different folds (one correct and one incorrect). Both folds are reflected in the data. In addition, a correct fold might be recreased and these additional correct recreased folds are reflected in the scoring as well. All proportions were calculated against the participants' total number of folds. Reaction time was not collected; our earlier studies of this task found it an ineffective predictor [c.f. 16].

From prior FoR studies, we predicted that the mobile device would lead to better user performance, since the opportunity for physical realignment potentially reduces the need for mental transformations. Specifically, based on prior research [7, 8, 12, 13] in which reducing mental transformations reduced errors, we predicted that the participants using the non-fixed mobile device would have the lowest PIF. We also predicted that PRF would be lessened for the mobile device participants because recreases could be indicative of an uncompleted and redone mental transformation.

3.1 Participants (Ps)

Our participants (Ps) included thirty-one graduate students in computer science classes. Ps were not distinguished by any background issues other than being students. Seven Ps used the fixed mobile device, thirteen Ps used the non-fixed mobile device, and eleven Ps used the fixed desktop to complete the task. All required ethical compliance guidelines regarding the treatment of Ps were followed throughout the study.

3.2 Materials and Task

The task for this study was to fold an origami whale, following the procedure of our prior studies. The procedure to create the whale consisted of 25 paper folds (and unfolds), with the instructions for making the folds presented in a series of 12 steps of varying degrees of complexity (Fig. 2).

Fig. 2. The completed origami whale

In order to control for variation in the presentation and also to accommodate the small screen size of the mobile device, the on-screen presentation consisted only of the interactive 3D presentation for each step (Fig. 3 - Right). Because a number of previous studies, including ours, suggest that Ps are better able to accomplish this task with a dual (verbal plus visual) presentation, we included a dual presentation as well [cf. 9]. Ps were given a paper packet with a text description of the step at the top and line drawing images of the step at the bottom. The line drawings included dashed lines and arrows that indicated what folds need to be made and showing what the object looks like at the beginning and end of each step. Each step was presented on a different page. The paper instructions were in a booklet that Ps could move. The paper instructions were printed on 8.5 x 11 paper; the line drawings occupied approximately half of each page.

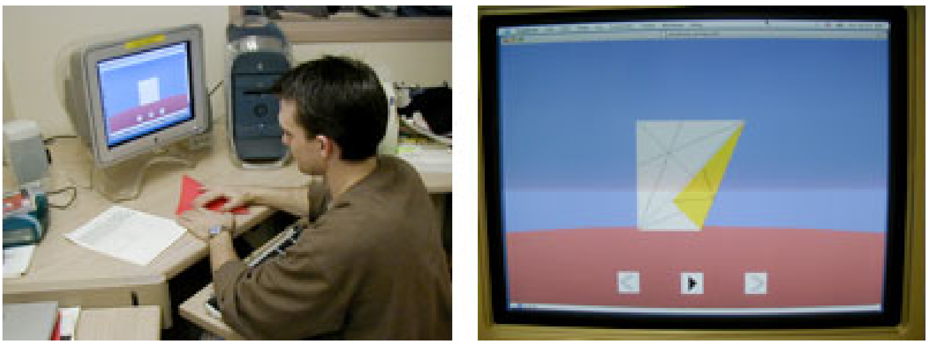


Fig. 3. The Desktop condition (left) and 3D interactive presentation (right)

3.3 User Interaction

Ps were permitted to move forward and backward through the presentation steps at their discretion and as many times as they wanted. There was no time limit for the task. Ps were shown both the paper instructions and the on-screen presentation but were not given any guidance as to how they should proceed. In other words, Ps were free to refer to either or both presentations as it best suited them. Ps could animate the presentation; they could stop the animation at any point and rotate the 3D model. The 3D interactive presentation was implemented using VRML 2.0 (Virtual Reality Modeling Language).

The desktop presentation was delivered on a Macintosh G4 running Mac OS 10.2 with 256M of memory. The VRML model was rendered within Internet Explorer 5.5 using the Cortona VRML client plug-in. The display device was a 17-inch Apple monitor with resolution set to 800x600 with 24-bit color. The monitor was set back approximately 18 inches from the user to provide desk space for the written instructions as well as space to fold the origami paper. Users used a standard one-button mouse. (Fig. 3 - Left).

The mobile 3D presentations were delivered with Microsoft Pocket PC version 3. The VRML model was rendered within Pocket Internet Explorer 5.5 using the Pocket PC Cortona VRML client plug-in. The display screen measured 3.8 inch (diagonal) with resolution 240x320 with 16-bit color. Additional desk space was provided for the participant to fold the origami paper. The Ps used a stylus to interact with the mobile device touchscreen. The mobile device was initially set on the desk in its cradle next to the written instructions. In the non-fixed condition, Ps could move the device to any position that they wished. (Fig. 4)

3.4 Procedure

Each participant received training on paper folding, 3D models, and the presentation interface. The training phase of the protocols took about 15 minutes per participant.

At the start of the actual experiment each participant received a piece of origami paper and instructions for their specific condition to fold an origami whale. A participant was considered to have completed the procedure and therefore had a scorable whale if the participant started the folding procedure and, eventually, he/she declared the task to be completed. The Ps' hands, the whale, and as much of the presentation on the screen as possible were videotaped.

4 Results

A significant multivariate analysis of variance (MANOVA) confirmed that the three dependent variables (PCF, PRF, PIF) were independent (Wilkes' Lambda = .675, $F(4.0, 54.0) = 2.935$, $p < .029$); thus we conducted separate univariate ANOVAs of the effect of the three conditions on all dependent variables. In one-way ANOVAs, only the PRF was significant ($F(2,28)=6.21$, $p < .006$); PCF and PIF were not significant. Pairwise, Ps with the fixed desktop display had a significantly higher PRF as compared to the other two conditions; while not significantly less, the PRF for the non-fixed mobile device was less than that of the fixed mobile device (means: fixed desktop display, 29%; fixed mobile device, 20%; non-fixed mobile device, 18%). The overall means and standard deviations of these variables are shown in Table 1.

Table 1. Overall means by category (standard deviations in parentheses)

Dependent Vars	Non-fixed Mobile (N=13)	Fixed Mobile (N=7)	Fixed Desktop (N=11)
PCF	70.80% (s.d. = 19.16)	74.51% (s.d. = 8.52)	59.96% (s.d. =14.50)
PRT	18.36% (s.d. =7.43)	19.64% (s.d. = 6.47)	28.91% (s.d.=8.72)
PIF	10.94% (s.d. = 17.40)	5.84% (s.d. =6.84)	11.12 (s.d. =11.63)

4.1 Qualitative Results

We viewed videos of the Ps, folding the whales in each of the three device conditions to determine if the Ps were physically aligning their folded whales to the presentation or vice versa. While this qualitative data does not reveal the number of mental alignments of FoRs, it is telling because it is 1) indicative that the Ps were using *a* strategy for realignment and 2) descriptive of how users may differentially utilize the features of the three device conditions to offload mental workload. In the case of the non-fixed mobile device, the participant could align the whale to the device and vice versa. In the other two fixed device cases, the participant could only align the whale to the device. All Ps could align the presentation itself with the whale or vice versa.

For the non-fixed mobile condition, initially the device was in the cradle. However, the Ps were given free reign to do whatever they wanted with the device itself. All but one took the device out of the cradle and proceeded to place it next to the paper that they were folding. Follow up analysis of video data indicated that 10 of the 13 Ps using the non-fixed mobile device physically moved the device and 5 of the 13 Ps clearly realigned the non-fixed mobile device to physically align the orientation of the displayed virtual object with the physical object they were creating.

For the Ps in the fixed mobile device, 4 of 7 Ps held the whale in alignment with the presentation. We observed that for this group, for those that did move the paper to the cradle, it was common to hold up the paper with one hand and then rotate the 3D whale model to align the two. For the Ps in the fixed desktop condition, 2 of 11 Pts aligned their whale with the presentation by holding the whale up to the presentation on the desktop screen or holding it in a similar orientation.

5 Discussion and Conclusions

While based only observational data, it is this last point, the relative alignment of the paper artifact with the computer display, that we conjecture to be an important aspect of the Ps' building process and their interaction with the presentation. In all three conditions, the Ps *could* have realigned the 3D presentation at any point and it is possible that the Ps in our study did this. However, we can verify that the Ps in the two mobile device conditions did realign at a high rate. The rate of realignment of the paper to the fixed non-mobile device was lower. This pattern likely explains the similarity between the two mobile device conditions. Ps in these conditions reduced their mental workload by moving the paper and/or the device.

Our results indicate that allowing the user to manipulate the device's position can influence performance by reducing the need for mental transformations. The reason why PIF was not significant while PRF was, unlike prior research that used context-free tasks with no clear correct answer presented to the participant [7, 12, 13], participants could see the correct end result for each step from the instructions. From the instructions, participants could see the correct fold for each step and they did not make errors per se. Participants who had more FoRs to realign recreased rather than erred. Perhaps the participants with the heaviest mental workload completed the task by approximation as a way to handle the heavier workload.

In all, our results suggest that participants using a mobile device were better able to complete our construction task with fewer errors compared to participants using a fixed desktop display. From our qualitative results, we can conclude that the advantage of the mobile device may lie in its ability to be rotated and aligned as required by the task at hand. Perhaps the ability to rotate and realign mobile devices to any angle as required by a task may suggest why mobile devices have become so popular and pervasive in the world.

Acknowledgments. We thank Ron Buchanan for his contributions to this project.

References

1. Barnes, J., Poor, G.M., Leventhal, L., Zimmerman, G., Klopfer, D.: Look and touch: The impact of touchscreens on the delivery of instructions for inherently 3D construction tasks using web-delivered virtual reality. Presented at the IPSI 2005 Conference, Amsterdam, The Netherlands (2005)
2. Cooper, L.A., Shepard, R.N.: Mental transformations in the identification of left and right hands. Journal of Experimental Psychology: Human Perception and Performance 104(1), 48–56 (1975)
3. Czerwinski, M., Tan, D.S., Robertson, G.G.: Women take a wider view. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves, pp. 195–202 (2002)
4. Eisenberg, M., Nishioka, A., Schreiner, M.E.: Helping users think in three dimensions: steps toward incorporating spatial cognition in user modelling. In: Proceedings of the 2nd International Conference on Intelligent User Interfaces IUI 1997, pp. 113–120 (1997)
5. Froehlich, P., Baillie, L., Simon, R. (2008). Realizing the vision of mobile spatial interaction. Interactions, 15-18 (January – February 2008)
6. Jolicoeur, P.: Orientation congruency effects on the identification of disoriented shapes. Journal of Experimental Psychology: Human Perception and Performance 16(2), 351–364 (1990)
7. Just, M.A., Carpenter, P.A.: Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychological Review 92, 137–172 (1985)
8. Klopfer, D., Athy, J., Leventhal, L.: Spatial ability and working memory: Just & Carpenter (1985) revisited. Poster session presented at the annual meeting of the Psychonomic Society, Long Beach, CA (2007)
9. Mayer, R.E.: Multimedia learning. Cambridge University Press, New York (2001)
10. Poor, G.M.: The effects of various levels of immersive technologies on a subject's ability to perform an inherently 3-dimensional construction task (Unpublished master's thesis). Bowling Green State University, Bowling Green, OH (2004)
11. Seager, W., Fraser, D.S.: Comparing physical, automatic and manual map rotation for pedestrian navigation. In: CHI 2007 Proceedings – Mobile Interaction Techniques II, San Jose, CA, USA, April 28-May 3, pp. 767–776 (2007)
12. Shepard, R.N., Hurwitz, S.: Upward direction, mental rotation, and the discrimination of left and right turns in maps. Cognition 18, 161–193 (1984)
13. Shiffrar, M.M., Shepard, R.N.: Comparison of cube rotations around axes inclined relative to the environment or to the cube. Journal of Experimental Psychology: Human Perception and Performance 17(1), 44–54 (1991)
14. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative effectiveness of augmented reality in object assembly. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 2003, pp. 73–80 (2003)
15. Waszak, F., Drewing, K., Mausfeld, R.: Viewer-external frames of reference in the mental transformation of 3-D objects. Perception & Psychophysics 67(7), 1269–1279 (2005)
16. Zimmerman, G., Barnes, J., Leventhal, L.: Delivering instructions for inherently-3D construction tasks: an evaluation of authoring environments for multimedia presentations. In: Proceedings of the 2000 International Computer Symposium, pp. 126–131 (2000)
17. Zimmerman, G., Barnes, J., Leventhal, L.: A comparison of the usability and effectiveness of web-based delivery of instructions for inherently-3D construction tasks on handheld and desktop computers. In: Proceeding of the Eighth international Conference on 3D Web Technology, Web3D 2003, pp. 49–54 (2003)

Part V

Development Methods, Algorithms, Tools and Environments

A Middleware Architecture for Designing TV-Based Adapted Applications for the Elderly

Carlos Gacimartín, José Alberto Hernández, and David Larrabeiti

Universidad Carlos III de Madrid
Avda. Universidad 30, 28911 Leganés, Madrid, Spain
{cgacimar,jahgutie,dlarra}@it.uc3m.es

Abstract. The elderly are begining to use more and more new technologies, although several times they find these are difficult to interact with, especially when they are accessed by a keyboard or the screen is too small.

This work proposes a middleware architecture for TV-based applications for the elderly, with a focus on social interaction services like webmail. The elderly feel very confident with their TVs since it is something they use on a daily basis, and the screen is sufficiently large. The architecture contains a Context Manager with an Ontology that models the elderly user impairments and actual context information collected from sensors located at the user's home, and an Interface Adaptation Engine which generates the adapted interface for a particular user, according to his profile and context of use. This architecture is part of an ongoing EU funded project: MyUI[1] and a first prototype is expected by mid 2011.

1 Introduction

In the last century, global life expectancy at birth has risen from 58 years in 1970-1975 to 67 years in 2005-2010, and is currently beyond 80 years for some developed countries, namely Japan (82.6 years, first ranked), Australia (81.2 years, fifth ranked) and Spain (80.9 years, sixth ranked). In developed countries, such an increase of longevity, together with the observed decrease of fertility, has a clear social impact: the elderly population is expected to nearly double by 2050 [1].

In light of this, Information and Communication Technologies (ICT) for the elderly are expected to increase their quality of life in the forthcoming years. The elderly are slowly embracing new technologies either because they are beginning to acknowledge their usefulness in daily tasks, but also because new technologies keep them close to the younger generations, especially their sons and grandsons. Also current generations which now demand technology will be technology consummers in the future. However, it is still the Television the preferred home appliance for the elderly. Indeed, it is the people beyond 65 who spend more

[1] http://www.myui.eu

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 443–449, 2011.
© Springer-Verlag Berlin Heidelberg 2011

time watching TV, with more than 294 minutes per day, 13 minutes more than the average TV consumption of people from all ages in Spain [2].

Hence, the combination of communication technologies and the Internet with TV will definitely open new services applications to people of all ages, and particularly to the elderly since the TV is the device they feel more confident with. A large number of TV vendors have forseen this opportunity and are producing TVs with Internet connection [3]. Additionally, several partnerships between TV vendors and Internet Service Providers have arisen to make true TV-based Internet services, see for instance Google TV with Sony and Logitech [4].

However, when developing ICT services for the elderly, a number of design requirements must be taken into account for its success, especially concerning accessibility issues in the Human Computer Interaction (HCI). Basically, the service presentation and navigation must be adapted to the aged user cognitive/visual/hearing profile of the user.

Indeed, elderly accessible software research is a hot research topic [6], see for instance Cybrarian [10], a webmail with a very easy-to-use interface intended for the elderly, but does not have interface adaptation. However, the use of a keyboard always hampers the adoption of ICT technologies by the elderly users. Ideally, a social interaction system for the elderly should be designed without the need to use a keyboard.

Take the webmail case for instance: A webemail without a keyboard is possible if the compose mail option is performed by recording the voice of the user. The user could navigate through the webmail on his TV and just press some buttons to record messages and send them to his family and friends. Additionally, font-size adaptation and even text-to-speech can be used for the hearing impaird users, where the emails font-size need to be increased or even read for the users.

In the following sections we propose a middleware architecture for the development of elderly-adapted applications, with a first focus on webmail applications, as part of the work conducted within the FP7 EU Project MyUI under code FP7-ICT-248606.

2 An Architecture for Services to the User

Fig. 1 shows a general system architecture for the design and implementation of web-based adapted services for elderly users. This architecture comprises three different parts: The Client side, the Context Manager, and the Middleware, which with a different role that is briefly described next.

2.1 The Client Side

The client side comprises the set of devices and sensors located near the user. These include:

- The Internet TV, which contains a web-browser that connects to the services allocated on the cloud. We assume that the elderly user accesses these services only with his TV remote controller.

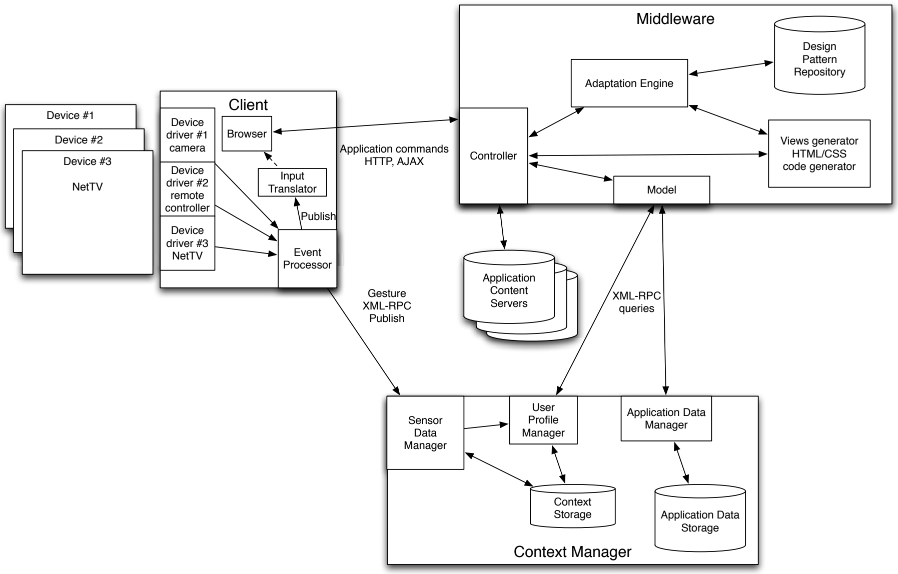


Fig. 1. System's architecture

- Hardware sensors, which are expected to capture real-time context information of the elderly user. Such information is first captured by the sensors' drivers and then published on the Context Manager using XML-RPC. Examples of important context information are: the user id that identifies who is accessing a given service (obtained from an RFID for instance), the user's distance to the TV, the current ambient light, etc. Such information is important in the generation of accessible interfaces.
- Software sensors, that are in charge of providing other inferred information about the user and his access to the service. For example, software sensors may measure the user's response time in accessing a given service or whether the user experiences difficulties in accessing the services. In summary, software sensors are expected to complement hardware sensors by inferring other contextual information about the user.

Hence, the client side not only must provide a front-end to access the adapted services, but also is expected to capture real-time context information about the user which helps in the generation of the personalised adapted interfaces. Such context information is first captured by the sensors, and then made available via XML-RPC on the Context Manager for further use by the Middleware.

2.2 The Context Manager

The Context Manager (CM) is the software component that stores both the user's profile and the measured context information. Both types of information shall be used by the Middleware and its Interface Adaptation Engine for the generation of adapted user interfaces.

The information concerning the user's profile and his current situation is modelled and stored in the CM by an ontology. This ontology defines, in a conceptual way, the type of information necessary to make adapted user interfaces. In the case of adapted services for the elderly, two types of information are modelled and stored in the ontology, which follows the Open Ambien Assited Living (AAL) framework [7] format:

- User profile, which contains personal information about the user, for example visual acuity, color blindness, hearing impairment, etc. This information is static and does not change over time.
- Context information, which contains information about the actual context of the user, for example, current distance to the screen, etc. This information is more dynamic, and it is updated with real-time measurements.

Typically, two operations are offered by the CM to the other software components in the architecture: setValue and getValue, the former allows for an update on the ontology's information whereas the latter permits information retrieval from the CM. Technically, both operations are served as Remote Procedure Calls (RPC) from external applications, and the information is embedded within XML (this is the XML-RPC communication paradigm).

The communication flow with the client side follows the asynchronous publish/subscribe paradigm, at which the sensors publish updated information on the CM. A triple is necessary to identify every sensor event, namely:

(sensor id, measured property, measured value for that property)

After this information is updated in the CM, then it is made available to other software components, for instance the Middleware. The Middleware retrieves both user profile and context information from the CM before generating the any adapted interface.

2.3 The Middleware

The Middleware provides both the application logic and the interface adaptation engine in the whole system. The Middleware is structured following the Model View Controllelr (MVC) [8] approach which separates a typical web-based application into specific areas:

- The Model or data layer. This layer contains the information required by the application, including access to databases and file-systems. In our application, the model interacts with the CM by quering for the static user profile and the more dynamic real-time context information related to that user.
- The Views or presentation layer. This layer is in charge of generating the final HTML and CSS code to be sent to the browser of the Internet TV. The presentation layer comprises the Interface Adaptation Engine, which translates design patterns and users profile and context information into HTML and CSS code customised for a given user.

- The Controller or logic layer. This layer represents the operation workflow, which typically comprises several steps:
  1. The Controller asks for the information required to the Model;
  2. this further queries the Context Manager which contains the updated information of the user and his context, which
  3. is sent back to the Controller. Then,
  4. this information is passed to the Interface Adaptation Engine which generates the adapted interface for that particular user and his context. This information arrives at the Controller which just
  5. forwards it to the web-browser in the client side. A summary of these steps is given in Fig. 2.

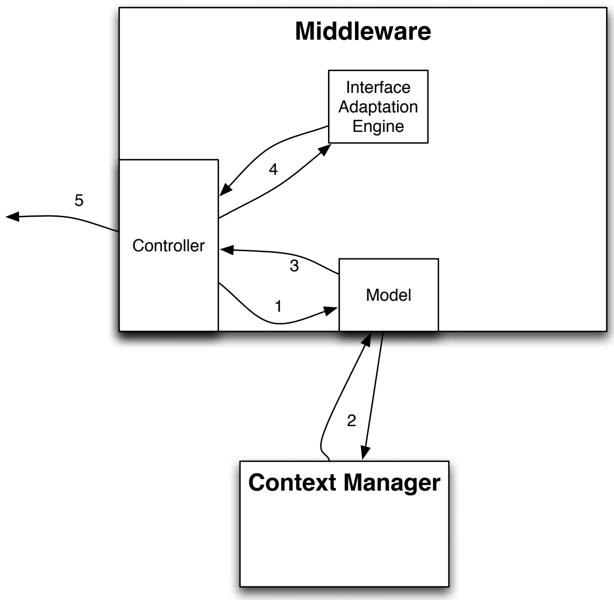


Fig. 2. MVC operation

Finally, the Controller also retrieves application-specific information from the Application Content Servers (Fig. 1). This refers to the contents of the application oriented for the elderly, for instance:

- Email application: The application contents in this case refers to the emails themselves.
- Physiotherapy exercise: In this case, the Controller needs to retrieve the flash videos that contain the specific exercises themselves.
- Cognitive games: The application information regards to the particular games along with their difficulty: Sudokus, Crosswords, Solitaire, etc.

Next, we examine how the CakePHP framework can be refined to provide such web-based TV services for the elderly.

3 Middleware Implementation with CakePHP

CakePHP is a native MVC framework that enables PHP developers to create robust web applications rapidly and reliably. CakePHP allows CRUD (Create, Read, Update and Delete) operations, necessary in most of web-based applications. By convention, urls in CakePHP always follows the same structure:

http://www.example.com/Application/Controller/Action/P1/P2/P3

which essentially calls some "Action" method from some "Controller" with some parameters "P1", "P2" and "P3". This url triggers all the steps in the Middleware: Contacts the CM and retrieves the user profile and context information, retrieves the application specific data and calls the Interface Adaptation engine with all this information to generate the HTML and CSS code adapted to the user.

Consider an email application, where an elderly user wants to read his emails after loging into the system with his username and password. In this case, the email controller would follow the next pseudo-code:

1. Identify the user, i.e. retrieve from the Context Manager using a XMLRPC call the RFID key of the user in front of the Internet TV. From that RFID code, the application has the necessary data for his session (login, email server, etc).
2. Retrieve the user's emails, namely, connect to the email server and retrieve the emails.
3. Text2Speech, that is, run some TTS code (for instance, the free Java-based TTS [9]) to convert the emails text to an audio format if necessary.
4. Dessign Patterns, in other words, apply a set of rules to translate the user's profile and his context into an appropriate interface with buttons, font-size, audio instructions, etc.

4 Summary and Conclusions

This work has proposed an architecture for a Middleware that provides adapted services to the elderly community. Essentially, the whole system contains:

1. An Internet TV and a set of sensors that measure real-time context information from the client.
2. A context manager, modelled by an ontology, which stores the user profile and the context information measured by the sensors.
3. A middleware platform that adapts the contents of some application with the context information obtained by the sensors.

The middleware follows the Model View Controller (MVC) approach which separates the data, presentation and logic layers. In our architecture, the Model refers to a set of functions that queries the Context Manager via XML-RPC, the

View comprises the Interface Adaptation Engine which generates HTML and CSS code from the model information and a set of pre-defined design patterns, and finally, the Controller handles all the operation workflow and logic, linking the application content, the model and the views. This architecture is currently under development within the EU-funded MyUI project, and a first version is expected to be released by mid-2011.

Acknowledgements

The authors would like to acknowledge the help of other partners from MyUI [5] in the development of this work, especially Matthias Peissner, Andreas Schuller and Dagmar Hbe from Fraunhofer IAO, Hester Bruikman and Roy van de Korput from Phillips Consumer Lifestyle, Peter Wolf and Oliver Strnad from Forschungszentrum Informatik, Ian Allen and Mark Ferguson from CleverCherry .com and Alberto Olmo, Víctor Sánchez and Alejandro García from ISOIN.

The authors would like to thank the support of the T2C2 Spanish project under code TIN2008-06739-C04-01 to the development of this work.

References

1. United Nations, Volume II: Sex and Age Distribution of the World Population "World Population Prospects, The 2008 Revision" (2008)
2. http://www.eldia.es/2011-01-03/sociedad/30-espanoles-vieron-television-media-minutos-persona-dia.htm
3. http://www.nettv.philips.com
4. http://www.google.com/tv
5. http://www.MyUI.eu
6. Sayaho, S., Blat, J.: Human-Computer Interaction with Older People: From Factors to Social Actors, Departament de Tecnologies de la Informaci i les Comunicacions, Universidad Pompeu Fabra (2009)
7. FZI Living Lab Ambient Assisted Living FZI Forschungszentrum Informatik, Germany
8. Rehman, K., Stajano, F., Coulouris, G.: An Architecture for Interactive Context-Aware Applications. Cambridge Syst. Associates (2007)
9. http://freetts.sourceforge.net
10. Dickinson, A., Newell, A.F., Smith, M.J., Hill, R.L.: Introducing the Internet to the over-60s: developing an e-mail system for novice Older computer users. Interacting with Computers, 1–22 (2005)
11. Age Concern, IT, Intemet and Older People, report on poll conducted for Age Concern by ICM Research, London, UK (July 2002), http://www.icmreseah.co.uk/reviews/2002/it-temet-old-peop1e.htm
12. Carmichael, A.R.: Style Guide for the Design of Interactive Television Services for Elderly Viewers, UK Independent Television Commission, London, UK (1999)
13. Hawthom, D.: Designing usable applications for older users an example, Conference, Melbourne, Australia (November 2002)
14. Rama, M.D., de Ridder, H., Bouma, H.: Technology generation and age in using layered user interfaces. Gerontechnology 1(1), 25–40 (2001)

Performance Visualization for Large-Scale Computing Systems: A Literature Review

Qin Gao[1], Xuhui Zhang[1], Pei-Luen Patrick Rau[1], Anthony A. Maciejewski[2], and Howard Jay Siegel[2,3]

[1] Department of Industrial Engineering, Tsinghua University, Beijing 100084, P.R. China
[2] Electrical and Computer Engineering Department
[3] Computer Science Department
Colorado State University, Fort Collins, CO 80523-1373 USA
{gaoqin,xuhuizhang,rpl}@tsinghua.edu.cn,
{aam,hj}@colostate.edu

Abstract. Recently the need for extreme scale computing solutions presents demands for powerful and easy to use performance visualization tools. This paper presents a review of existing research on performance visualization for large-scale systems. A general approach to performance visualization is introduced in relation to performance analysis, and issues that need to be addressed throughout the performance visualization process are summarized. Then visualization techniques from 21 performance visualization systems are reviewed and discussed, with the hope of shedding light on the design of visualization tools for ultra-large systems.

Keywords: performance visualization, performance monitoring, information visualization.

1 Introduction

Performance visualization is the use of graphical display techniques for the visual analysis of performance data. The process can be abstracted as a mapping from system behaviors to visual representations [1, 2]. By augmenting cognition with the human visual system's highly tuned ability to see patterns and trends, performance visualization tools are expected to aid comprehension of the dynamics, intricacies, and properties of program execution [3, 4]. Recently the need for extreme scale computing solutions [5] presents new challenges for the design and development of performance visualization tools. It is expected that next-generation supercomputers (so-called exascale computers) will be more than 1000 times faster than the current petascale systems. Such an increase in performance is associated with an increase in the level of parallelism, the heterogeneity of computing resources, and the complexity of interactions among these resources [6, 7]. Monitoring program run-time behaviors to tune performance in real-time is indispensible to realizing the potential of such systems, but the immense volume and complexity of the performance data presents a

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 450–460, 2011.
© Springer-Verlag Berlin Heidelberg 2011

major challenge for human perception and cognition. Powerful and easy to use performance visualization is important to solving this problem.

This paper presents a review of performance visualization tools for large-scale systems. The goal is not to exhaustively enumerate visualization systems, but to describe issues that need to be addressed throughout the performance visualization process, with the hope of shedding light on design of visualization tools for ultra-scale systems. In Section 2, a general approach to performance visualization is introduced, and issues related to each phase are discussed. Section 3 presents a review of visualization techniques from 21 performance visualization systems. The framework of information visualization by Card [3] is adopted to classify and analyze these techniques. Implications for future research are presented in Section 4.

2 Approach to Performance Visualization

According to literature on performance analysis and visualization [1, 3, 4, 11, 12], the process of performance visualization generally consists of four major steps: *instrumentation* (enabling access to performance data to be measured), *measurement* (recording selected data during the run-time of the program), *data analysis* (analyzing data for performance visualization), and *visualization* (mapping performance characteristics to proper visual representations and interactions).

A major challenge for instrumenting large systems is to decide what to be instrumented among a huge amount of data of different levels from numerous heterogeneous sources. The data to be collected should reflect application performance as closely as possible while minimizing perturbation of that behavior as much as possible (trade-off between fidelity and perturbation) [12]. The instrumentation may reside in hardware or in software. Hardware instrumentation involves a monitoring system collecting performance data, and generally incurs less performance degradation than software instrumentation. But the portability of this approach is low due to the requirement of dedicated monitoring hardware. Software instrumentation involves inserting small pieces of code, often referred to as *sensors*, in the operation system, the run-time environment, or the application program. This process can easily become tedious and time-consuming for large systems, and automation is desired to reduce the effort from application developers.

During the execution of applications, performance data can be measured by tracing (recording individual events and associated data) or by profiling (recording summary trends and statistics). Tracing provides more execution information, and is necessary for tools aiming to visualize detailed program run-time behaviors, such as [13-15]. Profiling collects only summary statistics, usually with hardware counters. This approach imposes less perturbation than tracing, but sacrifices fidelity. Tools using profiling, such as SvPablo [16], often allow data collection for large systems with long execution times. The recording action can be triggered either under specific conditions (event-driven, also called tracing by [11]) or periodically (sampling). Some tools support both approaches, such as [17]. The collected data can be analyzed and visualized during execution, as with Paradyn [18] and Virtue [13], or saved to a trace file for post mortem analysis, as with ParaGraph [2] and SvPablo [16]. For distributed applications in which resource availability changes dynamically, real-time measurement and visualization is necessary for effective performance tuning [13].

After a stream of performance data has been collected, data analysis will be performed to calculate various metrics required for performance evaluation and visualization, including microscopic metrics of individual components (e.g., processor state and execution rate) and macroscopic metrics of overall performance (e.g., concurrency, load balance). To extract useful information from a morass of performance data, data reduction methods (e.g., summation, averaging, extrema finding) are often used. Some systems make use of multivariate statistical analysis techniques, such as correlation, clustering, and multidimensional scaling, in search of recognizable relationships among related variables. In addition, a number of performance analysis systems developed application-specific analysis techniques for recognizing high-level program behaviors [19], pointing out causes of poor performance, generating scalability trends [9], and other application-specific purposes.

In the *visualization* phase, extracted performance metrics and relationships are mapped into selected visual structures, which are then integrated and transformed to form visualization views. Despite the seemingly hopeless variability of visual forms that could result, Card suggested that only a limited number of visual components are involved in information visualization due to various constraints to which such visualizations are subjected [3]. The basic visual components include the spatial substrate that defines the space, marks appearing in the space, connections and enclosures for linking and enveloping, retinal properties (graphic properties to which the retina of human eyes are sensitive independent of position, such as color, size, shape, texture, and orientation) overlaying other components, and temporal encoding for tracking changes of graphic properties.

3 Review of Visualization Techniques

3.1 Classification of Visualization Techniques

In this section, we will review visualization techniques from 21 performance visualization systems. These techniques can be classified into the categories of the information visualization taxonomy proposed by Card [3], as shown in Table 1.

3.2 Simple Visual Structures

Statistical charts with one or two variables. The simplest form of performance visualization is to use common statistical charts and diagrams, such as bar charts, pie charts, kiviat diagrams, and matrix views, to show summary statistics of computation, utilization, and communication metrics. These simple charts could be powerful, because they provide an overview of important performance metrics and enable quick identification of major problems, such as overload and imbalance. They are used widely in performance visualization systems, such as ParaGraph [2] and AIMS [9].

Table 1. Classification of performance visualization techniques

Category	Performance Visualization Techniques	Example applications and studies
Simple visual structures	Pie charts, distribution, box plots, kiviat diagrams	ParaGraph [2], PET [20], SvPablo [16], VAMPIR [21], Devise [22], AIMS [9]
	Timeline views	Paje [23], AIMS [9], Devise [22], AerialVision [24], Paraver [25], SIEVE [14], Virtue [13], utilization and algorithm timeline views in [17]
	Information typologies	SHMAP [26], Vista [4], Voyeur [27], processor and network port display in [28], hierarchical display in [12]
	Information landscape	Triva [29], Cichild [30]
	Trees & networks	Paradyn [18], Cone Trees [31], Virtue [13], [32]
Composed visual structures	Single-axis composition	AIMS [9], Vista [4]
	Double-axis composition	Devise [22], AerialVision [24]
	Case composition	Triva [29]
Interactive visual structure	Interaction through controls (data input, data transformation, visual mapping definition, view operations)	Paje[23], data input, filtering, and view manipulation in [28] and [32]
	Interaction through images (magnifying lens, cascading displays, linking and brushing, direct manipulation of views and objects)	Virtue [13], Cone Trees [31], Devise [22], direct manipulation of the 3D cone and virtual threads in [32]
Focus + context visual structures	Macro-micro composite view	Microscopic profile in [4], PC-Histogram in [24]

Timeline views. Use of timeline views in performance visualization systems can be classified into the following groups:

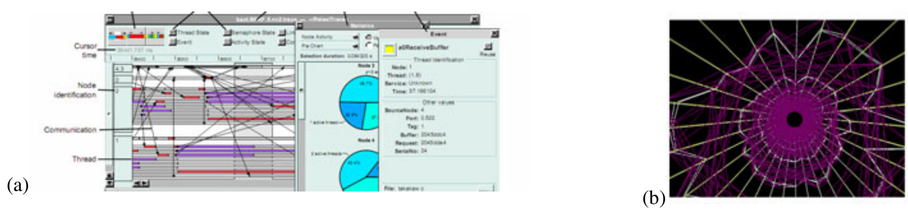


Fig. 1. Time views of program execution and communications. (a) Pajé: visualization of program execution and communication [33] (b)Virtue: time-tunnel display [13]

- *Describing run-time behaviors and communication paths*

Fig. 1(a) shows a typical visualization of program behaviors and communications. Time is mapped to the horizontal axis, and horizontal bars ranked along the vertical axis represent system components being analyzed, such as processors, tasks, and threads. Communications between the components are represented by arrows between bars. Colors are used to indicate the component status or the type of activities. The time lines can also be organized non-linearly, as in the time-tunnel display shown in Fig. 1(b). Other performance data can be added by using extra retinal features, such as textures and shapes [9, 15, 20]. A problem of such timeline views is that they scale poorly when system sizes increase. One way to mitigate the problem is to use more concise displays, such as the space-time diagram from ParaGraph [2]. Other methods to improve the scalability of such timeline views are discussed in [33, 34].

- *Showing the evolution of performance statistics over time*

This type of visualization plots summary performance metrics versus time, often using line plots (top plot in Fig. 2) or bar charts. It is often used to plot several different metrics or the same metrics related to different system components (e.g., different nodes, processes) versus time on a single view so they may be compared. Multiple metrics can be visualized as multiple line plots (middle plot in Fig. 2), stacked bar charts, or intensity plots (bottom plot in Fig. 2).

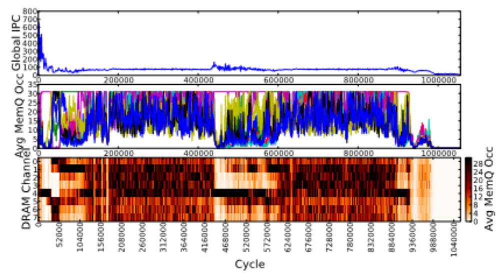
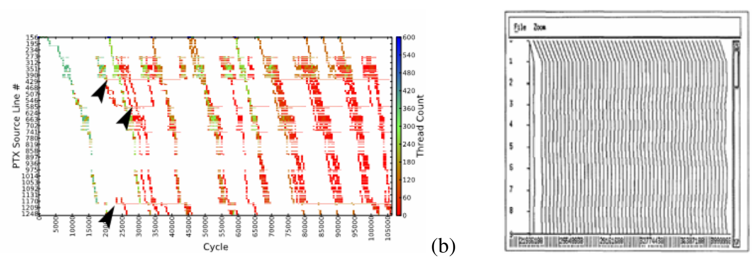


Fig. 2. Time views of utilization/computation/communication summaries of AerialVision, with the top plot showing global IPC, the middle plot showing the average memory queue occupancy for each DRAM channel, and the bottom plot showing the same data as the middle plot but using intensity plot (the darker the color, the higher memory queue occupancy) [24]



- (a) (b)

Fig. 3. Timeline views to facilitate source code level analysis. (a) AerialVision: PC-Histogram [24]. (b) SIEVE: Contour-plot [14]

- *Facilitating source code level analysis*

Fig. 3(a) presents a visualization of source code execution over time from AerialVision [24]. Program code is laid out in ascending source line number along the vertical axis. The time period in which the portion of a program is fetched in the pipelines is colored, where the intensity represents the number of threads fetching the portion of the program at the given time point. With this visualization, it is easy to identify the instructions that cause threads to become delayed, which are indicated by horizontal lines on the plot. Fig. 3(b) shows another visualization of source code execution from SIEVE [14], which displays calling to a specific function by processors as contours.

Information typography. Information typography is an essentially 2D scatterplot, with the two spatial dimensions partly defined by external structures, and retinal variables (e.g., colors, shapes) used as an overlay to add more information. Being able to combine performance data with physical topology, this visualization is especially useful for geographically distributed applications and systems with complex interconnection topologies. Fig. 4(a) shows an information typology of a large-scale cluster [28]. Network switches are arranged according to system interconnection topology. Colors are used to indicate the computational job and associated performance data, and thick line segments are used to show routing information once the route source and destination have been selected. Information typography can also be used to integrate logical architectures, as in [12].

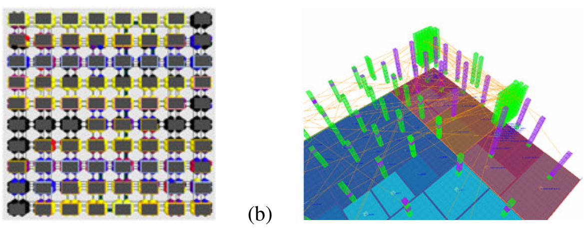
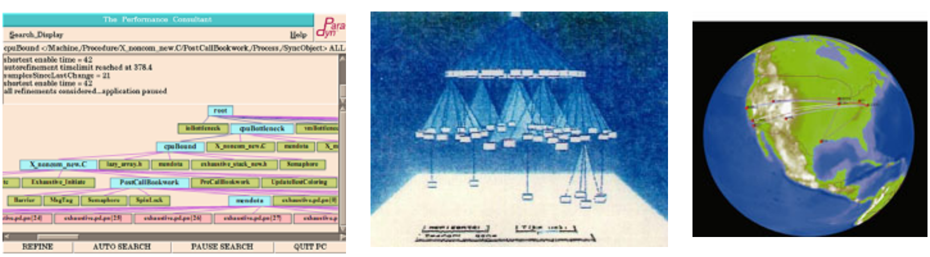

(a) (b)

Fig. 4. Information typography and information landscape. (a) Port display on network typology [28]. (b) Triva: information landscape based on resource hierarchy.

Information landscape. Information landscapes differ from information topographies in that they have one variable extruded into the third spatial dimension. Fig. 4(b) presents an information landscape from Triva [29]. The time axis is mapped to the vertical axis. Each vertical bar represents the progression of a process, and different states along the time axis are represented by different colors. These processes are put on top of a visual base (shown as colored rectangles), which is organized as a treemap [35] of the resource hierarchy.



(a) (b) (c)

Fig. 5. Trees and networks. (a) Paradyn: Performance Consultant, showing a search hierarchy [18]. (b) Cone Trees: 3D visualization of tree structures [31]. (c) Virtue: Geographic network display [13].

Trees and networks. Trees and networks differ from other visual structures in the use of explicitly drawn linkages for spatial positioning [3]. Trees are used to describe hierarchical information structures like resource hierarchies [36, 37]. AIMS uses trees to illustrate the dependency between hypotheses in a search for performance bottlenecks, as shown Fig. 5(a). A significant limitation of trees is that the nodes expand exponentially as a function of depth, which leads to an extreme aspect ratio for large trees. A possible solution is Cone Trees [31] (Fig. 5 (b)), which utilize 3D interactive visualization to make effective use of screen space and enable visualization of whole tree structures. Networks are useful for describing network connections, communication traffic, and relationships or communications among program entities. Fig. 5(c) shows a geographic network display from Virtue [13]. Vertices are placed on the global map according to the geographic distribution of resource sites, and retinal properties of vertices and edges are used to encode information like latency and bandwidth.

3.3 Composed Structures

A composed structure is a synthesis of two or more simple structures into a single compound view, with the aim to understand the relationship among variables or to provide a global image. There are several ways to produce a composed structure:

- Single-axis composition: multiple graphs sharing a single axis are aligned to produce a new view. Time, process, and resource identifiers are often used as the common axis. A special case of single-axis composition is *small multiples* [10], which presents a series of images showing the same combination of variables indexed by changes in another variable, and the information slices can be played as successive frames like a movie, or positioned to a third spatial dimension, as in [4].
- Double-axis composition: multiple graphs sharing two axes are plotted on a single graph. Fig. 6 shows an example from Devise [22]. The five small scatter plots show message bytes sent versus message bytes received for five processes, and the left plot is a composite of all of them, using different colors for encoding processes.
- Case composition: two different graphs describing the same cases can be fused into a single diagram by having a single mark for each case. Triva's information landscape shown in Fig. 6(b) is essentially a composition of a treemap of resource hierarchy and a communication timeline view.

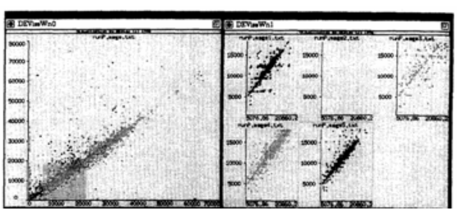


Fig. 6. Message traffic plot from Devise [22]

3.4 Interactive Visual Structures

Performance visualization systems provide a variety of user interactions to let the user select the portion of information they want to study and customize the way the information is presented, such as selecting alternative views, changing observation perspectives, zooming between different levels of detail, and manipulating other display parameters to acquire information they need to understand concurrent application behaviors. Interactive visual structures can be further divided into [1]:

- Direct interaction through the visualization: users manipulate the image directly to change the view. The simplest form of direct interaction is clicking a visual component to bring forward a more detailed view. An extension of this interaction is *magnifying lens*, which can be moved across the display and reveal hidden details when being hold over an object [13]. Other direct manipulations include panning the display by dragging, selecting visual objects, and re-positioning them. To make such direct manipulations more "natural" and immersive, Virtue [13] utilized virtual reality technologies to enable users to control the visualizations by gestures or direct manipulation of 3D sliders. For information at different grain sizes, *cascading displays* are often used, as in the case of Cone Tree shown in Fig. 5 (b). When two or more representations of the same data are presented together, the coordination between windows can be controlled through interactive visual structures, such as the graphical cursor in Devise [22].
- Indirect interaction though controls: users manipulate visualizations indirectly through predefined controls on the user interface. Interactions with the underlying computation, such as data-related controls (e.g., data input, filtering, grouping, and other transformation) and definitions of visual mappings (e.g., definition of colors, forms, and mapping relationships), are often carried out in this way. Common controls for view configurations include scroll-bars, zoom in/out buttons, sliders or input field for dynamic queries [28], and radio buttons for selecting specific characteristics to show [15].

3.5 Focus + Context Abstractions

Focus + context abstractions refer to visualizations whose mappings are modified by the computer, not by the users themselves, according to its model of the users' level of interest. In our survey, we found the use of such abstractions in performance visualization systems to be very limited. The only type being used is macro-micro composition [4, 24], in which both global picture and local details are provided, and detailed information accumulates into larger coherent structures. Focus + context abstractions are effective in combining overview and detail information in a single dynamic display. The scarcity of this type of visualization among performance visualization techniques may indicate directions for future performance visualization technique design.

4 Future Work

In this paper, we summarized issues that need to be addressed throughout the process of performance visualization and reviewed visualization techniques from 21 performance visualization systems using Card's framework of information visualization. These systems are built for monitoring and analyzing performance of parallel and distributed systems. Experiences gained and lessons learned from them could provide insight for the design of performance visualization tools for exascale computing systems, though there are new challenges. One major challenge is the huge size of exascale systems, which requires good scalability of performance visualization techniques. On the one hand, data abstraction methods that are used widely in scientific visualization may be adapted to performance visualization. On the other hand, visualizations based on focus + context abstractions may help to combine high-level context and low-level details in a more effective way. Other challenges are ergonomics and usability issues related to performance visualization for super-scale systems. Without considering the limitations of human sensory and cognition capabilities, the complexity of performance visualizations may easily exceed them. These issues should be addressed by future research in the field of visualization.

Acknowledgment. This research was supported in part by the US National Science Foundation Grant number 0905399, and the CSU Abell Endowment.

References

1. Roman, G.C., Cox, K.C.: A taxonomy of program visualization systems. Computer 26, 11–24 (2002)
2. Heath, M., Etheridge, J.: Visualizing the performance of parallel programs. IEEE Software 8, 29–39 (1991)
3. Card, S.: Information visualization. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 509–543 (2002)
4. Rover, D.T.: A performance visualization paradigm for data parallel computing. In: Twenty-Fifth Hawaii Int. Conf. System Sciences, pp. 149–160 (2002)
5. Simon, H., Zacharia, H., Steven, R.: Modeling and simulation at the exascale for energy and the environment. Technical report, Department of Energy (2007)
6. Shestak, V., Smith, J., Maciejewski, A.A., Siegel, H.J.: Stochastic robustness metric and its use for static resource allocations. Journal of Parallel and Distributed Computing 68, 1157–1173 (2008)
7. Ali, S., Maciejewski, A.A., Siegel, H.J., Kim, J.K.: Measuring the robustness of a resource allocation. IEEE Trans. Parallel and Distributed Systems 15, 630–641 (2004)
8. Miller, B.P.: What to draw? When to draw? An essay on parallel program visualization. Journal of Parallel and Distributed Computing 18, 265–269 (1993)
9. Yan, J., Sarukkai, S., Mehra, P.: Performance measurement, visualization and modeling of parallel and distributed programs using the AIMS toolkit. Software Practice & Experience 25, 429–461 (1995)
10. Tufte, E.R.: Envisioning Information. Graphics Press, Cheshire (1990)
11. Kraemer, E., Stasko, J.T.: The visualization of parallel systems: An overview. Journal of Parallel and Distributed Computing 18, 105–117 (1993)
12. Koehler, S., Curreri, J., George, A.D.: Performance analysis challenges and framework for high-performance reconfigurable computing. Parallel Computing 34, 217–230 (2008)
13. Shaffer, E., Reed, D.A., Whitmore, S., Schaeffer, B.: Virtue: Performance visualization of parallel and distributed applications. Computer 32, 44–51 (2002)

14. Sarukkai, S.R., Gannon, D.: Parallel program visualization using SIEVE.1. In: 6th Int. Conf. Supercomputing - ICS 1992, pp. 157–166 (1992)
15. De Kergommeaux, J.C., Stein, B., Bernard, P.E.: Pajé, an interactive visualization tool for tuning multi-threaded parallel applications. Parallel Computing 26, 1253–1274 (2000)
16. De Rose, L., Reed, D.: SvPablo: A multi-language architecture-independent performance analysis system. In: 1999 Int. Conf. Parallel Processing, pp. 311–318 (1999)
17. Alexandrov, A., Armstrong, D., Rajic, H., Voss, M., Hayes, D.: High-level performance modeling of task-based algorithms. In: Int. Symp. Performance Analysis of Systems and Software (ISPASS), pp. 184–193 (2010)
18. Miller, B.P., Callaghan, M.D., Cargille, J.M., Hollingsworth, J.K., Irvin, R.B., Karavanic, K.L., Kunchithapadam, K., Newhall, T.: The Paradyn parallel performance measurement tool. Computer 28, 37–46 (1995)
19. Bates, P.C.: Debugging heterogeneous distributed systems using event-based models of behavior. ACM Trans. Computer Systems (TOCS) 13, 1–31 (1995)
20. Garcia, J., Entrialgo, J., Garcia, D.F., Diaz, J.L., Suarez, F.J.: PET, a software monitoring toolkit for performance analysis of parallel embedded applications. Journal of Systems Architecture 48, 221–235 (2003)
21. Nagel, W.E., Arnold, A., Weber, M., Nagel, W.E., Arnold, A., Weber, M., Solchenbach, K.: VAMPIR: Visualization and Analysis of MPI Resources. Supercomputer 12, 69–80 (1996)
22. Karavanic, K.L., Myllymaki, J., Livny, M., Miller, B.P.: Integrated visualization of parallel program performance data. Parallel Computing 23, 181–198 (1997)
23. De Kergommeaux, C., Stein, B.D.O.: Flexible performance visualization of parallel and distributed applications. Future Generation Computer Systems 19, 735–747 (2003)
24. Ariel, A., Fung, W.W., Turner, A.E., Aamodt, T.M.: Visualizing complex dynamics in many-core accelerator architectures. In: Int. Symp. Performance Analysis of Systems and Software (ISPASS), pp. 164–174 (2010)
25. Pillet, V., Labarta, J., Cortes, T., Girona, S.: Paraver: A tool to visualize and analyze parallel code. Transputer and Occam Developments, WOTUG-18, pp. 17–31 (1995)
26. Dongarra, J., Brewer, O., Kohl, J.A., Fineberg, S.: A tool to aid in the design, implementation, and understanding of matrix algorithms for parallel processors. Journal of Parallel and Distributed Computing 9, 185–202 (1990)
27. Socha, D., Bailey, M.L., Notkin, D.: Voyeur: Graphical views of parallel programs. 1988 ACM SIGPLAN and SIGOPS Workshop on Parallel and Distributed Debugging - PADD 1988. pp. 206–215 (1988)
28. Haynes, R., Crossno, P., Russell, E.: A visualization tool for analyzing cluster performance data. In: 2001 IEEE Int. Conf. Cluster Computing (CLUSTER 2001), pp. 295–302 (2001)
29. Schnorr, L.M., Huard, G., Navaux, P.O.: Triva: Interactive 3D visualization for performance analysis of parallel applications. Future Generation Computer Systems 26, 348–358 (2010)
30. Brown, J.A., McGregor, A.J., Braun, H.W.: Network performance visualization: Insight through animation. In: Passive and Active Measurement Workshop, pp. 33–41 (2002)
31. Robertson, G.G., Mackinlay, J.D., Card, S.K.: Cone Trees: Animated 3D visualizations of hierarchical information. In: SIGCHI Conf. Human Factors in Computing Systems: Reaching Through Technology, pp. 189–194 (1991)
32. Kim, Y.J., Lim, J.S., Jun, Y.K.: Scalable thread visualization for debugging data races in OpenMP programs. In: Cérin, C., Li, K.-C. (eds.) GPC 2007. LNCS, vol. 4459, pp. 310–321. Springer, Heidelberg (2007)

33. De Kergommeaux, J., De Oliveira Stein, B.: Pajé: an extensible environment for visualizing multi-threaded programs executions. In: 6th Int. Euro-Par Conf. Parallel Processing, pp. 133–140 (2000)
34. Knüpfer, A., Voigt, B., Nagel, W., Mix, H.: Visualization of repetitive patterns in event traces. In: Workshop State-of-the-Art in Scientific & Parallel Computing, pp. 430–439 (2007)
35. Johnson, B., Shneiderman, B.: Tree-Maps: A space-filling approach to the visualization of hierarchical information structures. In: 2nd Conf. Visualization 1991, pp. 284–291 (1991)
36. Heisig, S.: Treemaps for workload visualization. IEEE Computer Graphics and Applications 23, 60–67 (2003)
37. Goldberg, J.H., Helfman, J.I.: Enterprise network monitoring using treemaps. In: 49th Annual Meeting of the Human Factors and Ergonomics Society, pp. 671–675 (2005)

Developing a User Recommendation Engine on Twitter Using Estimated Latent Topics

Hiroyuki Koga[1] and Tadahiro Taniguchi[2]

[1] Graduate School of Engineering, Ritsumeikan University, Japan
[2] Department of Human & Computer Intelligence, Ritsumeikan University, Japan
koga@em.ci.ritsumei.ac.jp

Abstract. In recent years, microblogging is popular among people and informal communication becomes important in various communities. Therefore, a number of Web communication tools are developed to facilitate informal communication. In this paper, focusing on microblogging service, Twitter, we develop a user recommendation engine which extracts latent topics of users based on followings, lists, mentions and RTs. This recommendation algorithm is based on Latent Dirichlet Allocation (LDA) and KL divergence between two users' latent topics. This algorithm hypothesizes that the users have latent connection if the distance calculated by KL divergence is short. Additionally, we performed an experiment to evaluate the effectiveness of the algorithm, and this showed that there is correlation between the distance and user's preference obtained through questionnaire.

Keywords: LDA, Twitter, Information Recommendation.

1 Introduction

Informal communication is quite important in various communities because that is not only for fun but also connects a person to other people providing with important information. For example, some researchers insist that communication in a smoking room accelerates participants to share information. Then, many studies have been made on informal communication which is beneficial for their daily works and facilitates information sharing [1,2,3]. These researches don't limit their offline communication on inquiry. To facilitate online informal communication is, rather, a hot research area.

Web communication services become widely used and developed. To the services, like message boards, microblogging and Web chat, everyone can access, read and write in public equally; that is, personal relationships between users had not been included in the services. However, in recent years, services which focus on the personal relationships become more popular; users only accesses, reads and writes other users' pages based on their own private relationships. After the facebook [4] emerges, Web communication tools which treat personal informations and the relationships become popular. Twitter [5] is one of the most popular tools among them. To obtain fruitful information on twitter, a user should find and follow other users who often tweet messages in which he/she is interested. He/she can read only the messages of them

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 461–470, 2011.
© Springer-Verlag Berlin Heidelberg 2011

he/she follows. The messages tweeted by them he/she follows are shown on his/her Time Line (TL). This means that to find users who are suitable for him/her is important on this informal communication architecture. In this paper, we develop a user recommendation algorithm by using Latent Dirichlet Allocation (LDA). LDA is a well-known document clustering method [6]. Usually, LDA is adapted to documents whose elements are words. In contrast, we apply this to a set of followings of a user and other information on Twitter. We also evaluate effectiveness of this recommendation algorithm through an experiment.

2 Background

2.1 Informal Communication Support

Recently, many researchers are studying on informal communication support system. Nakano et al. proposed the Traveling Café to support informal communication in a local community [1]. The system support informal communication between coffee drinkers by prompting person who make a cup of coffee in a coffee room to go to pour another's cup of coffee. A system proposed by Siio et al. helps people interact with each other by assuming a place around a coffee machine as informal communication space [2]. These studies tried to trigger interaction in the real world using ICT.

On the other hand, new online communication services including SNS and microblogging become booming. In these services, users can communicate with the users' friends of the real world naturally on the Web, because users can edit friends and community based on their relationships. In this paper, we focus on informal communication on Twitter. Our goal is to encourage users to enjoy informal communication on Twitter. Therefore, we develop user recommendation engine recommending users who share the similar interest with a target user. In our proposed method, latent topics which represent users' interests are estimated statistically by using LDA.

2.1 Twitter

Twitter is a Web service on which we can talk to each other with short messages, which called "tweets". The length of a tweet is limited to 140 characters. Thus, the feature of this service is that users can send messages other users more frequently than Web chat or mail. In this service, "following" and "follower" are keywords[1]. If a user follows other users, their tweets are displayed on his/her Time Line (TL). Tweets of users he/she does not follow are not displayed on his/her TL. Therefore, communication is hardly generated based on their tweets. So, if the number of followings is a little, he/she would not use Twitter effectively. In contrast, if he/she follows many users randomly, they may tweet in which he/she is not interested. His/her TL is filled with their uninteresting tweets because their tweets are displayed on his/her TL. So that, communication on Twitter becomes worthless. Therefore it is important for users to find and follow users who post interesting tweets for them. However, on Twitter, we try to find our friends or interesting people who share similar topics with us, keyword search used

[1] We name the users he/she follows as his/her followings and name the users who follow him/her as his/her followers.

frequently on Web is not effective method, because the messages on Twitter is quite shorter than they on Web pages.

Currently, users usually use alternative methods. The methods include browsing other users' followings, followers, RT, lists and mentions. RT is a user's tweet by using another user's tweet (Fig.1). List is a grouping function of a user's followings. He/she can group his/her followings by using this function. Mention is his/her tweet to a specific user (Fig.2). These functions give hyperlinks for him/her to find referential users. We presuppose that the user and the user's followings, users who retweets messages which generate interests and users who mention for the user have common interest. Though, there are problems for followings and lists. If we can find a complete list which follows entire users we want to follow, only we have to do is simply to follow the list. However, a list is always customized just for one user who made the list. A user always customizes a list for his/her self. So, we always cannot find a complete list for us. List is usually edited from one user's personal point of view for personal use. In the same way, the users cannot make a complete set of their followings.

In this paper, we extract statistically latent topics that generate followings and lists by using probabilistic model. We assume that users who share same latent topics with a user are worth following for him/her. We also assume that topics are represented by clusters of human relations and categories by users on Twitter.

Fig. 1. Example of RT: kogame5 tweets by using tweet of tryal

Fig. 2. Example of mention: tanichu tweets to kogame5

3 Latent Topics Extraction

In this section, we introduce Latent Dirichlet Allocation (LDA) [6] to extract latent topics which generate followings and lists in a sense of generative model. It was developed to cluster for document, originally. It generates clusters of words and documents based on Bayesian statistical model. It is a generative model that assumes a document is multinomial distribution of topic z which generates word w. Our algorithm clusters users on Twitter using LDA by considering a user as a document. A user's followings' names, names of lists that include the user, names of target users whom the user mentioned, and labels of RT are considered as words contained in a document. The labels of RT are user's name who tweets the original message and number which was put in time-series when the original message was tweeted. Fig.3 represents LDA's graphical model. α and β is hyperparameter. We use four kinds of Twitter user information to make corpus for using LDA. These are described briefly in the next section.

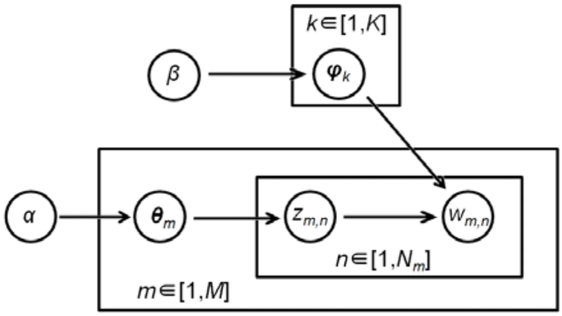


Fig. 3. Graphical model of LDA

3.1 Types of Corpora

We made four types of corpora using four kinds of user information; following, RT, mention and list. Our algorithm extracts latent topics by applying LDA to this corpus.

1. following corpus
2. RT corpus
3. mention corpus
4. list corpus

Following corpus is a set of names of users whom the user follows. It also contains his/her names. *RT corpus* is a set of labels which consist of a user's name who tweets the original message and number which was put in time-series when the original message was tweeted; that is, we do not use text of RT[2]. *Mention corpus* is a set of names of users to whom a user tweets addressing. *List corpus* is a set of names of list which contains the target user and it was made by the other users. We made these four kinds of corpora. By combining some of them, a training data set for a user is prepared. Additionally, we appended a meaningless string to each users' dataset to avoid each user's dataset to be empty. In our experiment, we used Twitter API to get Twitter users' data [7].

3.2 Latent Dirichlet Allocation (LDA) [6]

LDA is a method to estimate latent topics outputting each words by hypothesizing that document is constructed by words which are generated based on *k* topic. Griffiths proposed the method [8] which sample from a mixture model with multiple multinomial distributions based on Gibbs Sampling [9,11,12] with LDA. Thus, using this method, we can sample around global optimal solution without local solution. Distribution (1) can be obtained by a probabilistic argument [8,12].

---

[2] There are two kinds of RT that are official and non-official. We used non-official RT.

$$P(z_i = k \mid z_{-i}, w) \propto \frac{n_{k,-i}^{(t)} + \beta}{\sum_{t=1}^{V} n_{k,-i}^{(t)} + V\beta} \frac{n_{m,-i}^{(k)} + \alpha}{\sum_{k=1}^{K} n_{m}^{(k)} + K\alpha} \tag{1}$$

where, z_i is the i th topic; w is a set of words; $n_{k,-i}^{(t)}$ is the number of t th word appeared in the k th topic excluding i th topic; $n_{m,-i}^{(k)}$ is the number of k th topic appeared in the m th document excluding i th document; α is the hyperparameter for θ_m which is a parameter of a distribution which generates probability of the m th document; β is the hyperparameter for φ_k which is a parameter of a distribution which generates probability of the k th topic; V is number of words; K is number of topics; M is number of documents. First right-hand member is probability of i th word in k th topic, second right-hand member is probability of k th topic in m th document.

Gibbs sampling algorithm for latent Dirichlet allocation
while not finished do
 for all documents $m \in [1, M]$ do
  for all words $n \in [1, N_m]$ in document m do
   for the current assignment of k to a term t for word $w_{m,n}$:
  decrement counts and sums: $n_m^{(k)} - 1$; $n_m - 1$; $n_k - 1$;
   multinomial sampling acc. to Eq.(1) (decrements from previous step):
  sample topic index $\tilde{k} \sim p(z_i \mid z_{-i}, w)$
   use the new assignment of $z_{m,n}$ to the term t for word $w_{m,n}$ to:
  increment counts and sums: $n_m^{(\tilde{k})} + 1$; $n_m + 1$; $n_{\tilde{k}}^{(t)} + 1$; $n_{\tilde{k}} + 1$
  end for
 end for
end while

Each step of the Gibbs sampling procedure involves replacing the value of one of the variables by a value drawn from the distribution of that variable conditioned on the values of the remaining variables. That is, z_i is replaced by the extracted value from distribution $f(z_i \mid z_{-i}, w)$. This procedure is repeated either by cycling through the variables in some particular order or by choosing the variable to be updated at each step at random from some distribution. Here, M is the number of data, T is the number of step.

3.3 Recommendation Method

Our method calculates distance between two users by comparing the users' probabilistic latent topics. It measures the distance between users' multinomial distributions representing topics they are interested in with KL divergence. It selects potential users recommended to a target user if the KL divergence is small enough. KL divergence from the target user P to a recommended user Q is defined as bellow.

$$D_{KL}(P \| Q) = \sum_{x=0}^{T} p(x)\{\log(p(x)) - \log(q(x))\} \tag{2}$$

where, x is the topic index, T is the number of topics, $p(x)$ is a generation probability of the x th topic of a user P. The x th topic is generated from a user P. $q(x)$ is a generation probability of the x th topic of a user Q. A user Q whose distance $D_{KL}(P \| Q) < 1.5$ [3] and whom user P doesn't follow is recommended to user P.

4 Experiment

4.1 Experimental Condition

We extracted RT data and mention data from 50 users' TLs data from 2010/6/28 to 2010/7/1, following data and list data from profiles of 50 users selected for an experiment at 2010/7/1. We prepared corpora by mixing following corpus, list corpus, RT corpus and mention corpus. We didn't use RT corpus alone because the amount of RT data is not enough to construct multinomial distributions. Similarity, we didn't use mention corpus alone, either. By using these data, the LDA constructed multinomial distributions of topics for each users and multinomial distribution of words for each topic.Fig.4 indicates an example multinomial distribution of one user.

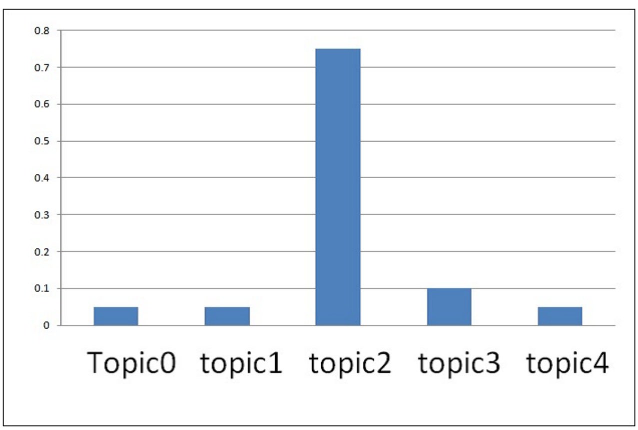


Fig. 4. Example of multinomial distribution of user's topic

[3] Here, 1.5 is decided with heuristic.

We assume that this distribution corresponds to the target user's interests. This figure indicates generation probability of user's topics. For example, this figure shows that topic 2 is generated with the probability of 75%. Our algorithm recommends users based on the KL divergence between two users, we conducted experiments to evaluate effectivity of our proposed method. We selected eight participants from fifty users, we showed TL of recommended candidates to participants. The candidates include *near users*, *distant users* and *half distant users*. The near users have distance of $D_{KL}(P \| Q) < 1.5$. The distant users are up to 2 users of maximum of $D_{KL}(P \| Q)$. The half distant users are up to 3 users of half maximum of $D_{KL}(P \| Q)$ around. We got eight participants to answer three questionnaires after looking through the TLs, followings and lists of the candidates.

Table.1 shows kinds of questionnaires.

Table 1. The three kinds of questionnaires for participants

	Questionnaire	Answer
Questionnaire 1	Do you want to follow this user?	1. I want to follow this user. 2. I may follow this user. 3. I do not want to follow this user.
Questionnaire 2	Are you acquaintance with this user?	1. I am acquaintance with this user. 2. I am not acquaintance with this user.
Questionnaire 3	Had you followed this user?	1. Yes, I had. 2. No, I had not.

We got participants to answer questionnaire 1 “Do you want to follow this user” on a scale of 1 to 3 that are (1)I want to follow this user (2)I may follow this user (3)I don't want to follow this user. We got participants to answer questionnaire 2 “Are you acquaintance with this user” on a scale of 1 to 2 that are (1)I am acquaintance with this user (2)I am not acquaintance with this user. Questionnaire 1 and 2 are asked to evaluate efficiency of this recommendation algorithm. Besides, Twitter users often remove their followings. If a target user had followed and removed some recommended users once, he/she would not want to follow them again. However, we could not get the history, the result contains unrelated factor which we want to take out. Therefore, we prepared questionnaire 3 to know the recommended users are followed or not followed. In this experiment, participants looked twenty four users in average. We prepared four kinds of corpora. Table.2 shows four types of corpora. We conducted an experiment and questionnaire each experimental conditions. In this experiment, we used LDA program “LDA implementation in C++ using Gibbs Sampling” developed and provided by Phan [10].

Table 2. The four kinds of corpora in experiment

Condition names	Kind of corpus
C1	Following
C2	List
C3	Following + list
C4	Following+list+RT+mention

4.2 Experiment a Result

In this section, we show results of the experiments. Fig.5 shows the results of questionnaire 1, this indicates the average of KL divergence with error bars representing standard deviations each experimental conditions. Fig.5 shows distance between multinomial distributions correlate to averages of questionnaire 1's answer. More in detail, participants have tend to answer (1)"I want to follow this user" or (2)"I may follow this user" if the distance between participant and a recommended user is small enough. Fig.6 shows the number of the target users' answer of questionnaire 1 to the candidates in each condition. On $C1$, $C3$ and $C4$, we took similar results. Answer rate of (1)"I want to follow this user" is slightly higher on $C3$ than $C1$ (46% on $C3$, 40% on $C1$). The number of candidates tends to be decreasing on $C1$ because participants

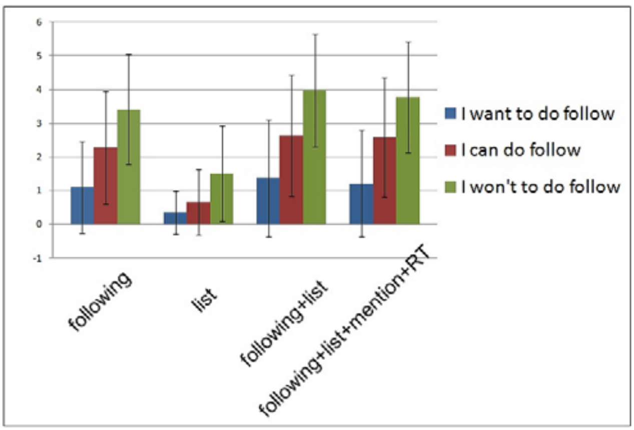


Fig. 5. Relation between result of questionnaire and KL divergence each experimental conditions

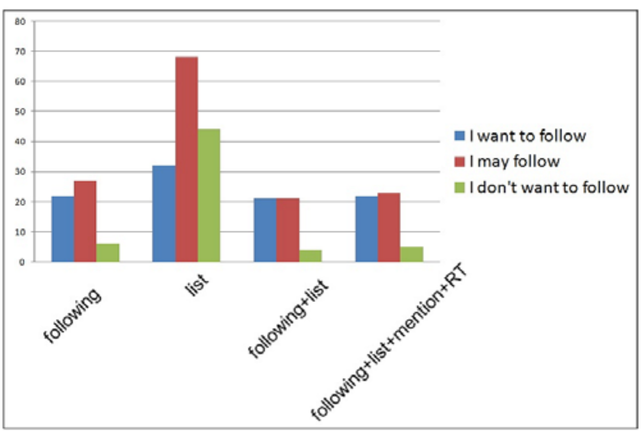


Fig. 6. Number of answer of questionnaire 1 each experimental condition

already followed the candidates. We found that it is not necessarily that recommendation result is interior division blending a number of corpora. This result indicates that a method to blend some kinds of data is important. Skipping detailed result, rate of answer (1) on $C1$ (27%) and on $C2$ (11%) vary greatly. This result indicates that there is qualitative difference between recommended users on each condition. On the other hand, lists are made by a user editing his/her followings. Therefore, the lists are not simple subsets. We skip explaining about questionnaire 3 because there is not discriminative result.

5 Conclusion

In a document clustering with LDA, Blei focuses on a relation between a document and words in the document. In this paper, we replaced the relation with a relation between a user and his/her followings or lists. We extracted multinomial distributions of users' topics by replacing relations between words and documents in LDA with between a user and the user's followings and lists. We found out similarities of users by comparing distance of multinomial distributions of users' topics. However, users' RTs and mentions data that are active information of users are not valuable. The reason why RTs and mentions are not valuable is that an experiment term is short. We obtained good result by recommending users who have a multinomial distribution that is similar to a multinomial distribution of a target user. We treated user's information which has quantitative difference between the user's followings and lists as same quality things. On another front, it is important to combine effectively these informations by considering difference of quality of these informations. To combine effectively these informations, we refer to Multimodal LDA [13] which is proposed by Nakamura et al. Multimodal LDA that treats a number of information as same quantitative information, for concept acquisition.

References

1. Nakano, T., Kamewada, K., Sugito, J., Nagaoka, Y., Ogura, K., Nishimoto, K.: The Traveling Cafe: A Communication Encouraging System for Partitioned Offices. In: Conf. on Human Factors in Computing Systems (CHI 2006) (April 2006)
2. Siio, I., Mima, N.: Meeting Pot: Coffee Aroma Transmitter. In: UbiComp 2001: International Conference on Ubiquitous Computing. ACM, Atlanta (2001)
3. Matsubara, T., Sugiyama, K., Nishimoto, K.: Raison D'etre Object: A Cyber-Hearth That Catalyzes Face-to-face Informal Communication. In: Han, Y., Tai, S., Wikarski, D. (eds.) EDCIS 2002. LNCS, vol. 2480, pp. 537–546. Springer, Heidelberg (2002)
4. facebook, `http://www.ja-jp.facebook.com/`
5. Twitter, `http://www.twitter.com/`
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation The Journal of Machine Learning Research (2003)
7. Twitter API Wiki,
`http://www.apiwiki.twitter.com/w/page/22554648/FrontPagetwitterAPI20.txt`

8. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences 101 (suppl. 1), 5228–5235 (2004)
9. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis, 2nd edn. Chapman & Hall CRC, Boca Raton (2003)
10. Phan, X.H., Nguyen, C.T.: LDA implementation in C++ using Gibbs Sampling, `http://gibbslda.sourceforge.net/`
11. Bishop, C.M.: Pattern Recognition And Machine Learning, pp. 542–546. Springer Science+Business Media, LLC, Heidelberg (2006)
12. Heinrich, G.: Parameter estimation for text analysis. Technical report (2004)
13. Nakamura, T., Nagai, T., Iwahashi, N.: Grounding of word meanings in multimodal concepts using LDA. In: Proc. IEEE/RSJ International Conference on Robotics and Automation 2009, pp. 3943–3948 (October 2009)

Project and Development of ErgoCoIn Version 2.0

Marcelo Morandini, Roberto Leite de Moraes Rodrigues,
Marcus Vinicius Cerrato, and Marcos Lordello Chaim

School of Arts, Science and Humanities, State University of Sao Paulo, Av. Arlindo Bettio, 1000, Ermelino Matarazzo, Sao Paulo, SP, Brazil, 03828-000
{m.morandini,roberto.rodrigues,mcerrato,chaim}@usp.br

Abstract. Usability evaluation (UE) methods may have several purposes: identify and diagnose usability problems; evaluate the implementation, comparing the one under evaluation with other systems, and also perform acceptance test. These methods are usually expensive, time-consuming and complex. In this context, the ErgoCoIn 1.0 has being used as an accessible system that can be used by several people all over the world, even if they are not usability experts or evaluators. This paper aims to present a new version of this environment considering that its basic objective is to propose a higher quality for the questions. Additionally, a new presentation questionnaire window small enough to be used both in web browsers and mobile devices is proposed.

Keywords: Usability Evaluation, Checklist, Automation.

1 Introduction

Software developers and Human-Computer Interaction (HCI) designers always deal with a common problem: evaluate the usability of the systems they developed. Specifically, concerning the internet based systems, that are usually under constant evolution, their features can be analyzed as the former *web sites* [4].

Usability, as pointed by the ISO/IEC 9241, can be defined as: the *effectiveness*, *efficiency*, and *satisfaction* with which users achieve specified goals in particular environments. Effectiveness can be defined as the accuracy and completeness with which specified users can achieve specified goals in particular environments. Efficiency as the resources expended in relation to the accuracy and completeness of goals achieved. And, finally, user satisfaction as the comfort and acceptability of the work system to its users and other people affected by its use [2].

Specifically, some web sites usability evaluation techniques [1; 2; 3; 6; 7; 8; 10] use previous evaluations approaches that produce qualitative reports, which might lead to subjectivity problems. This paper presents an objective evaluation approach that aims to quantify the efficiency, efficacy and user satisfaction while interacting with a web site to perform pre-defined tasks, and this approach is based on the server log files registers. We wish to confirm that it is possible to go further in the log files analysis in order to present usability studies. Using the log files data mining, it is possible to obtain taxes and metrics that quantify the usability of some tasks, producing trustworthy, objective, fast and low cost evaluations.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 471–479, 2011.
© Springer-Verlag Berlin Heidelberg 2011

A basic demand of any website is that its web application should work to fulfill its intended purpose – this is addressed by research into Web Engineering methods [1, 13, 14], models and tools. However, there is also the equally important demand that the web application must be usable by the visitors of the website. Usability research includes work on sets of guidelines which help to improve website usability. The ErgoCoIn environment [12] is an approach designed to give support to non-ergonomic expert evaluators in such a way they could conduct contextual, objective and systematic usability evaluations of web sites and their pages. It postulates the identification of the context of use aspects, comprising: users, tasks and environment attributes concerning the web site under evaluation. Content of the interviews/questionnaires [11] and the other environment data gathering activities is based on information demand underlined by the base of ergonomic questions initially compiled in this project. Such a strategy allows for proceeding very objective ergonomic inspections. Only pertinent information are proposed to inspectors in the contextual analysis and only applicable questions are presented to them in the web site inspection. Objectiveness is obtained also by inspections questions that are linked to a logic construction based on ergonomic criteria [9] that define their relative importance.

2 Web Site's Usability Evaluation

The usability of a product is affected not only by the features of the product itself, but also by the characteristics of the users, the tasks they are carrying out, and the technical, organizational and physical environment in which the product is used.

The main quality in an User Interface (UI) is its usability, which establishes how much the systems projected are easy to use and learn [1, 6]. Usability Evaluation (UE) methods have several purposes, such as: to aid the Web UI project to reach the user needs; to identify and diagnosis problems; and to perform a cognitive usability evaluation (comparing with other systems, and also for the acceptance test). The collected data during these UEs can be qualitative (descriptions that characterize the usability) or quantitative (measured that quantify the usability). Some interactive systems UE methods found in literature [1; 4; 5, 6, 10] approaches UI qualitative aspects; however they are not enough objective to quantify the efficiency, effectiveness and user satisfaction.

Therefore, UE is an important part of the overall user interface design process, which consists of iterative cycles of designing, prototyping, and evaluating [8]. Usability evaluation is itself a process that entails many activities depending on the method employed [13]. Common activities include: *Capture* collecting usability data, such as task completion time, errors, guideline violations, and subjective ratings; *Analysis* interpreting usability data to identify usability problems in the interface; and *Critique*: suggesting solutions or improvements to mitigate problems [4].

A wide range of usability evaluation techniques have been proposed. Some evaluation techniques, such as formal user testing, can only be applied after the interface design or prototype has been implemented. Others, such as heuristic evaluation, can be applied in the early stages of design. Each technique has its own requirements, and generally different techniques can uncover different usability problems. Usability findings can vary widely when different evaluators study the same user interface, even if they use the same evaluation technique [1,4,10].

Automation of usability evaluation has several potential advantages such as the following:

- Reducing the cost of usability evaluation.
- Increasing consistency of the errors uncovered.
- Predicting time and error costs across an entire design.
- Reducing the need for evaluation expertise among individual evaluators.
- Increasing the coverage of evaluated features.
- Enabling comparisons between alternative designs.
- Incorporating evaluation within the design phase of UI development, as opposed to being applied after implementation.

It is well known that the average quality of websites is poor, "lack of navigability" being the main cause of user dissatisfaction [10]. On the one hand web technologies evolve extremely fast, enabling sophisticated tools to be deployed and complex interactions to take place. Secondly, the life cycle of a website is also extremely fast: maintenance of a website is performed at a rate that is higher than that of other software products because of market pressure and lack of distribution barriers. In addition, often the scope of maintenance becomes so wide that a complete redesign takes place.

On the other hand, the quality of a website is rooted on its usability, which usually results from the adoption of user-centered development and evaluation approaches [1, 10, 12, 13, 14]. Usability testing is thus a necessary and repeated step during the life-cycle of a website. To test usability of a website a developer can adopt two kinds of methods: usability inspection methods (e.g. heuristic evaluation [1, 5, 13] or user testing [9]). Heuristic evaluation is based on a pool of experts that inspect and use a (part of a) website and identify usability problems that they assume will affect end users. With *user testing*, a sample of the user population of the website is selected and is asked to use (part of the) website and report things that they think did not work or are not appropriate.

Even though the cost (in terms of time and effort) of both methods is not particularly high, and their application improves the website quality and reduce the overall development cost, they are not systematically performed at detailed levels on every different part of a website after each maintenance or development step.

Specifically, the checklists can be applied by design staff, not necessarily experts in usability evaluation, such as: programmers and users. The checklists are extremely cheap to be proposed and only need some few users to be effective. They can also be virtually proposed at any time in the development life cycle, i.e., from the very initial HCI prototype untill a concluded and under use web site [11, 14].

3 The ErgoCoIn Version 1.0

Web sites development became accessible to a great amount of designers, not just the experts on usability or ergonomics fields. So, if the developments are always faster, the products developed can present a lack of quality [15]. According to this, we have defined an approach for the ErgoCoIn environment considering that there is the need that non-expert designers should be presented with a minimal quantity of information concerning a strategy for their designs as the context of use must be well defined and understood. A second consideration is that web sites are often designed through a fast and low cost design processes supported by tools which lead designers to carry out numerous and, sometimes obvious, ergonomic flaws.

ErgoCoIn is a checklist-based environment initially aimed at supporting objective ergonomic inspections of e-commerce web site and pages [16]. The ErgoCoIn features automatic inquiring services to identify context of use aspects (users and environment attributes) and to recognize web page components. Consequently, it is able to propose to inspectors only questions applied to the specific task context of use and to the associated web page components [12]. The ErgoCoIn validation phase was based on case studies composed by repetitive cycles of inspections to be carried out by a group of inspectors. The idea is to verify objectiveness and systemic attributes like results repeatability and reproducibility at the end of each cycle. In the last cycle, problematic questions can be deleted. These case studies will also be used for the analysis of the inspection tasks and for the specification of the most fitted user interface for support this task. ErgoCoIn logic architecture can be filled in by inspection questions and criteria concerning the different kinds of web sites application domains. Initially, the original knowledge content is supposed to be specific to e-commerce, due to its importance and also to the high number of web sites in this domain that can be used to validate this environment.

Aiming the ErgoCoIn´s definition, two evaluation techniques have been selected: Inspection Methods and Questionnaires. Inspection Methods do not require the direct participation of users (as it is easier and less costly) [9, 12]. They regroup a set of approaches that use the evaluators' judgment, being experts or not in usability. Even if these various methods may have different goals, they generally aim to detect the interface features that can lead to inefficiencies or higher workload for the users. The inspection methods distinguish themselves by the way the evaluators' judgment is performed and by the evaluation criteria that are the basis for the judgments. Specifically, we selected inspection techniques that are strongly supported by heuristics or even better, by individual recommendations.

The questionnaires allow not only to collect subjective data related to the users' actions and comments about their satisfaction from web sites, but also about the achievement of factual elements concerned to the system under evaluation [11]. The interviews/questionnaires with users can be used to identify features of the context of use, the usability difficulties, and collect local meaningful users' comments. They are even considered as a way to establish comparisons between the scenarios prescribed by the designer and the "real" ones, as described by the user.

The design of the usability evaluation technique underlying the ErgoCoIn has been motivated by four considerations: (1) that web sites development became accessible (through easily available design tools) to a large spectrum of "designers", not necessarily highly skilled in computer science or in ergonomics; (2) web sites are often designed along a fast and low cost design process supported by non expensive tools which lead designers to carry out numerous and sometimes obvious ergonomic flaws; (3) the high cost of usability evaluations using the "traditional" methods; and (4) the subjectivity that is a common problem in any usability evaluation.

The constraints listed above lead to the main characteristic of the intended method: an inspection-based evaluation with directly accessible web-based support mechanisms (strategies, recommendations, etc.), and information on context of use coming from users and designers. Several parameters must be set when selecting components for an evaluation method.

Therefore, it was defined that all this knowledge should be directly collected from the users and also from the designers using questionnaires or interviews for this purpose. So, due to this definition, the approach only can be applied to real web sites that are being used regularly and it is necessary to have the designers and some users available for the interviews or, at least, able to answer some questionnaires [11].

ErgoCoIn combines the Inspection Method and Questionnaires Evaluation techniques in an approach able to allow rapid and context focused ergonomic inspections [4, 5, 12]. The inspection component resulting from examining a large collection of ergonomic recommendations later completed with other data collected from different studies to elaborate checklists for the ergonomic characteristics applicable on web sites [8, 9]. These recommendations were formulated as questions and associated with both an ergonomic criterion that allows defining a system of relative importance between questions. Interviews/questionnaires and guidance for collecting data from users and designers were defined from analyzing the information demands in each question we elaborated. ErgoCoIn approach's is divided into two main phases: web site's Contextual Analysis and its Evaluative Inspection, as presented in Figure 1.

The ErgoCoIn´s Contextual Analysis phase collects all information related to the web site operational contexts that are useful for the usability evaluation process. This phase consists of a site description process and interviews of the users and designers. Also, ErgoCoIn supports and automatizes these activities. A HTML component recognizing tool identifies the existence of specific web site user interface components and aspects and organizes them according to two categories of descriptions: the global web site and the individual web pages associated with the main tasks accomplishment [12].

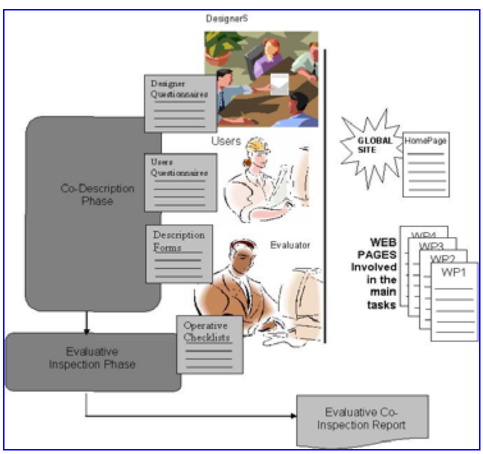


Fig. 1. The ErgoCoIn approach framework [12]

The second phase of the method is formed exclusively by inspection activities. ErgoCoIn starts the process performing an automatic analytical evaluation based on the comparison between user´s and designer's information concerning the intended and the real context of use features. The system points out to existence of designer's misconceptions about user´s features, and indicates the web site aspects to verify or reformulate in consequence.

At that, the system assembles checklists concerning the overall site and the specific web pages features related to task scenarios. These checklists can be considered as "objective" ones, once they will propose only the site components applicable questions arranged according to their levels of importance [12]. Applicability decisions result from processing the site description stored in the context of use database. Priority decisions results from ranking the Ergonomic Criteria [9] according to usage context features. A default Ergonomic Criteria ranking is suggested, but it can be modified by the evaluators, according to the characteristics of the current web site context of use. In fact, the original importance structure was proposed with a general Business to Consumer usage context in mind, in which non professionals users operate sites of virtual stores from their home environments aiming to buy simple products in a relatively low frequent basis. In such a situation the Guidance criterion should be considered before the Workload criterion. Anyway, ErgoCoin can authorize evaluators changing the importance structure at the ergonomic criteria level to accommodate different usage contexts.

4 The ErgoCoIn Version 2.0

This section aims to present a new version of the ErgoCoIn environment considering that its basic objective is to propose a higher quality for the questions that should be applied in the questionnaires that may be proposed. This new questionnaire proposition allows the definition and realization of objective ergonomic inspections by gathering pertinent information concerning the web site´s supported tasks, proposing only pertinent questions that should be presented to the inspectors that are relevant and significant and can be answered during the web site´s inspection. So, among others, we are updating the following activities:

1. A presentation questionnaire window evolution that now allows the filling of the answers in a smaller and second level window. This window only has one visible question per time and has visible interactions buttons presented, such as next and previous questions and help. This new small window will always be visible to the ErgoCoIn´s users and it may be possible to move it over the screen in order to allow the inspectors to interact with the web site to answer the question. Also, we are proposing the dimensions for this new window to be small enough to be used both by web browsers and mobile devices.
2. The questions that compose the questionnaire can be updated (deleted, inserted or modified) by the evaluator, even if these questions may belong to any context group or are related to any specific ergonomic criterion [9]. This aims to avoid "Not Applicable" answers for the questions.
3. Each one of the questions is now related to a specific Ergonomic Criterion. This is important for future analysis, which includes a quantitative ranking calculation based on the criteria.

As in the version 1.0, the evaluative inspections are performed by an evaluator applying the set of checklists defined in the previous phase. As mentioned before, this process constitutes an evaluative inspection as the evaluator is asked to judge the quality of very precise web site features. The questions in this method could not be answered by an intelligent or expert system once they require consideration of several factors not always at the reach of such systems. However, the level of judgment proposed by questions was defined in accordance with the level of ergonomic knowledge expected from evaluators (fairly basic usability expertise). Indeed, the questions phrases and associated support information, like justification and examples, were formulated in order to be easily understandable. Even so, this method leads to results which are subjective in nature and which may not point at all major ergonomics problems. As mentioned earlier, it is viewed as a rapid and widely usable mean to detect usability design faults, but in which human involvement is required.

Figure 2 presents the description window for the questions that compose the questionnaire to be applied. Although this window is still written in Portuguese, it is easy to realize that there are important information presented:

1. a specification of the main ergonomic criterion;
2. the presentation of the question itself;
3. a description that explains the question, including important information, such as references and links; and
4. the kind of the answer that is divided in two possibilities – Yes/No or a leveled answer, that represents the possibility of given a score as an answer, such a 0 – 10 ranking.

Fig. 2. Question Presentation Window

Also, Figure 3 presents a window that presents the ranking of the Ergonomic Criteria according to a definition proposed by the evaluators and also according to the ranking of the answers presented by the ErgoCoIn´s users.

We believe that using this new version we can get more efficacy and efficiency on web sites´ usability evaluation. Therefore, the reports generated may present meaningful data to designers, what include usability taxes and metrics that may help designing new products or improving the ones that are already being used and have being evaluated.

We have performed some validation tests of this new version, comparing it with the application of "traditional" usability methods and also with the version 1.0. The results obtained indicate that this new approach is more attractive as it presents more conclusive and meaningful reports.

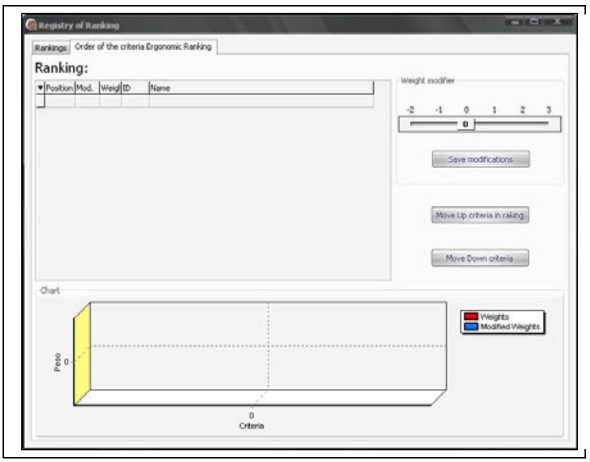


Fig. 3. Screen Aimed at Receiving Definitions Concerning Relative Ergonomic Criteria Importance

Now, to validate this new version of ErgoCoIn we must:

1. define a group of web sites to be evaluated. They must belong to different application domains and need to have specific well defined tasks to be evaluated (ie, tasks that have initial and final web pages well defined);
2. define a group of evaluators, aiming to have different skills and capabilities when interacting with the web sites under evaluation;
3. prepare and perform several usability evaluations in order to collect data to calculate usability taxes and metrics;
4. perform traditional usability evaluations of these web sites; and
5. compare the results obtained by both using the EgoCoIn version 2.0 and the traditional usability evaluation methodologies.

These steps are proposed as future works to improve and validate the ErgoCoIn version 2.0.

References

1. Downton, A.: Engineering the Human-Computer Interface. McGraw-Hill, London (1992)
2. ISO/DIS 9241. Dialogue Principles in Guiding the Evaluation of User Interfaces, Guidance on Usability (1997)
3. Brajnik, G.: Automatic Web Usability Evaluation: What Needs to be Done? In: 6th Conference on Human Factors and the Web, Austin, Texas, USA (2000)
4. Ivory, M.Y., Heasrstam, M.A.: The State of the Art in Automating Usability Evaluation of User Interfaces. ACM Computing Surveys 33(4) (December 2001)
5. Nielsen, J.: Designing Web Usability, Editora Campus, p. 416 (2000)
6. Storrs, G.: The Notion of Task un Human-Computer Interaction. In: HCI 1995- 10th Annual conference of the British Human-Computer interaction Group. University of Huddersfield, UK (1995)
7. Winckler, M., Palanque, P., Farenc, C., Pimenta, M.: Task-Based Assessment of Web Navigation Design. In: Proc. of TAMODIA 2002, Task Models and Diagrams for User Interface Design, Bucharest, Romania (2002)
8. Scapin, D.L., Leulier, C., Vanderbonckt, J., Mariage, C., Bastien, C., Palanque, P., Farenc, C., Bastilde, R.: Towards Automated Testing of Web Usability Guidelines. In: Tools for Working with Guidelines, pp. 293–304. Springer, London (2001)
9. Scapin, D.L., Bastien, J.M.C.: Ergonomic Criteria for Evaluating the Ergonomic Quality of Interactive Systems. Behaviour and Information Technology 16(4/5) (1997)
10. Molich, R., Bevan, N., Curson, I., Butler, S., Kindlund, E., Miller, D., Kirakowski, J.: Comparative Evaluation of Usability Tests. In: Proceedings of the Proceedings of the Usability Professional's Association Conference (1998)
11. Wammi: Website Analysis and MeasureMent Inventory (Web Usability Questionnaire), (n.d.) (2005), `http://www.ucc.ie/hfrg/questionnaires/wammi` (accessed in 2009)
12. Cybis, W.A.: ErgoManager: a UIMS for monitoring and revising user interfaces for Web sites . Rocquencourt: Institut National de Recherche en Informatique et en Automatique (2005) (Research report)
13. Keynote, A. Research Manager Clickstream, `http://www.keynote.com/solutions/solutions_cem_research_mgr_clickstream.html` (accessed in 2010)
14. Sommerville, I.: Software Engineering. Prentice-Hall, Englewood Cliffs (2009)
15. Morandini, M., Cybis, W.A.: Definition and Prototyping of ErgoMonitor. In: Proccedings of International Conference on Human-Computer Interaction, Creta, vol. II, p. 1396–1399 (2003)
16. Cybis, W.A.: UseMonitor: suivre l'évolution de l'utilisabilité des sites web à partir de l'analyse des fichiers de journalisation. In: 18eme Conférence Francophone sur l'Interaction Humain-Machine, Montréal. Actes de la 18eme Conférence Francophone sur l'Interaction Humain-Machine, vol. 1, pp. 295–296. ACM - The Association for Computing Machinery, New York (2006)

A Reference Model for Adaptive Visualization Systems

Kawa Nazemi, Christian Stab, and Arjan Kuijper

Fraunhofer Institute for Computer Graphics Research,
Fraunhoferstrasse 5, 64283 Darmstadt, Germany
{kawa.nazemi,christian.stab,arjan.kuijper}@igd.fraunhofer.de

Abstract. One key issue of both Information Visualization as well as Adaptive User Interfaces is information overload. While both disciplines have already devised well performing algorithms, methods and applications, a real merging has not taken place yet. Only a few attempts bring the surplus values of both disciplines together, whereas a fine-grained investigation of visualization parameterization is not investigated. Today's systems focus either on the adaptation of visualization types or the parameterization of visualizations. This paper presents a reference Model for Adaptive Visualization Systems (MAVS) that allows the adaptation of both the visualization type and the visualization parameterization. Based on this model, a framework for the adaptive visualization of semantics data will be derived. A use case describing the interaction with an "adaptive visualization cockpit" covering different visualization metaphors concludes the paper.

Keywords: Adaptive Visualization, Information Visualization, Intelligent Visualization, Visualization Reference Model, Ontology Visualization.

1 Introduction

Information overload is a well-known phenomenon in the current information age [1]. Different research areas face this phenomenon with their own methods and try to reduce the interaction cost [2], cognitive overload [3] and information amount, using content- and collaborative-based recommendation techniques. The community of intelligent and adaptive systems investigates this problem with novel and promising algorithms and techniques for user-, content- and activity analysis and adaptation. On the other hand, the Information Visualization (IV) community develops appropriate techniques for presenting information graphically.

For the community of IV, the way from data-oriented visualization to a more human-centered information presentation plays a key role. In 2007, one of the ten main challenges for IV was the inclusion of semantics or contexts [4]. In the following year, especially the human as an implication and decision factor for IV was placed in the foreground of the research [1]. The increased involvement of user's intentions and preferences in the forming process of IV is also noticeable in the challenges and scopes of Visual Analytics in 2008, where the adaptation of IV systems was proclaimed. Thus, one of the most important challenges is the development of "novel interaction algorithms incorporating machine recognition of the actual user intent and

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 480–489, 2011.
© Springer-Verlag Berlin Heidelberg 2011

appropriate adaptation of main display parameters such as the level of detail, data selection, etc. by which the data is presented" ([5], p. 162).

Researchers in the field of Adaptive Systems [6] represent a similar position and are increasingly recognizing the importance of the usage of visualization techniques [7]. This research area disposes of a comprehensive pool of methods, systems and algorithms for recognizing and analyzing user related information. With these methods Adaptive User Interfaces (AUIs) facilitate the handling with complex information and support users during their work process [8]. Different existing systems e.g. intelligent help systems [9], personalization of web page navigations [10] or command line proposal lists [11] are already using these methods and tailor the user interface to the given usage context. These methods and systems do mainly not refer to bridge the gap between IVs and AUIs. Actually only a few attempts try to bring the areas of IV and AUIs together [12]. These approaches focus either on the adaptation of the visualization type, using a single set of parameters [13] or on the adaptation of visual parameters of a single visualization type [7]. Despite the research in both areas, a system or a reference model for adaptation of visualization types and their visual parameters could not be found. Further the aspect of composing multiple visualizations [14] in adaptive systems is not investigated yet.

In this paper we introduce a reference Model for Adaptive Visualization Systems (MAVS) that provides the opportunity to adapt both the visualization type as well as visual parameters. Beside a detailed description of the reference model we present an implementation of the MAVS as a framework. Finally, we introduce an application example that not only shows the adaptation based on an example but also opens new research questions for the adaptation of visualizations.

The remaining paper is structured as followed: In the next section, the related work is presented in three sections: Reference models, frameworks and applications. Afterward the reference model is described followed by a detailed description of its implementation. An application example of the implemented framework concludes the description of our work.

2 Related Work

The related work is sub-divided in three parts: 1) Reference Models will describe existing work in an abstract level. 2) Frameworks will describe existing work on the implementation of the reference models and 3) Applications will describe existing systems that combine Adaptive Systems and Information Visualizations.

From the IV point of view Card et al. [15] have presented a general visualization reference model, which has been applied successfully to different visualizations. This model contains different phases that are needed for mapping and transforming raw data into a visual representation. The user may control the way in which the data is transformed, or mapped to visual structures. So the reference model focuses on the creation of visualizations and does not investigate any kind of adaptations. Based on the established reference model of Card et al. different extensions exist today [5,16], which investigate different issues of visual data representation more in depth. An extension with the investigation of a visual adaptation is only investigated by Aaltonen and Lehikoinen [17], who redefined the data table and visual structures of the original model. They incorporated rules into the data table and renamed it as context table. In addition the visual structure was divided into layers (overlapping visual representations). The main difference is the usage of rules for the context. These rules consist of simple structured logical inferences and logical- and comparison operators. The aspect of a user-centered or task-based adaptation in a context-free environment (non-mobile) is not investigated.

The Information Visualization community provides a large set of frameworks for visualizing information in different ways. The most famous is Prefuse introduced by Heer et al. [18], a framework for creating dynamic visualizations of structured and unstructured data, based on the visualization reference model of Card et al. [15]. All existing frameworks support the development and creation of visualizations and are often more a programming platform than a basis for adaptive visualizations. The issue of automatic or semi-automatic adaptation is not investigated. The only framework, we could find, which considers the aspect of adaptation is e-Viz [19]. Whereas e-Viz addresses the modeling, scheduling and managing of visualization computational tasks and rendering graphics. e-Vis does not provide a framework or reference model for adaptation of visualization types and visualization parameters to meet the users' demands.

To round out the related work concrete applications will be introduced in this part. Applications are the most concrete alternatives to the described reference models and frameworks and conclude the related work section. Already different approaches have been proposed to support an automatic or semi-automatic adaptation of UIs. Cicero, for example is a component-oriented architecture [19], where a central UI adaptation manager is used. On the contrary agent-based environments [21], where UI models are transformed and rendered into different platforms, adapt user interfaces to different impact factors. The only applications that adapt the visualization to certain impact factors, e.g. user interactions or user-goals are the following applications: Gotz and Wen [13] analyze user interactions and extract behavioral-patterns that are used to recommend different visualization types (Line Graph, Fan Lens, Parallel Coordinates or Bar Charts) to the user. But a visualization type has always characteristics and parameters, e.g. color of entities, order, size, layout etc., which can and should be used for communicating designated information adaptively. On the contrary Ahn & Brusilovsky are adapting visual parameters of a single visualization type and visualize the user-specific relevance of a query [7]. In this case a single static visualization type is used to represent the searched information.

The related work could point out that the number of existing adaptive visualizations is very limited and these approaches focus either on the adaptation of the visualization type, using a single set of parameters or on the adaptation of the parameters of a given single visualization type. Despite the research in both areas, a system or a reference model for adaptation of visualization types and their parameters could not be found. Further the aspect of composing multiple visualizations [14] in adaptive systems is not investigated yet.

3 A Reference Model for Adaptive Visualization Systems

In this section, we present a Reference Model for Adaptive Visualization Systems *MAVS*. The MAVS (Fig. 1) consists of three basic Components: *Input*, *Adaptation* and *Output*.

These Components are composed by several *modules*, which are defined by interfaces and show a unique behavior to their Component. Internally, there can be different conformant implementations, allowing a very easy reuse. In the reference model, we will assume that there is a toolbox of implementations available at runtime, allowing the adaptation engine to select the best fitting one for a certain situation. Additionally these modules can be configured by *parameters*. For instance these parameters are defined by certain impact and influence factor, e.g. knowledge or behavior of the user or the structure and amount of the underlying data.

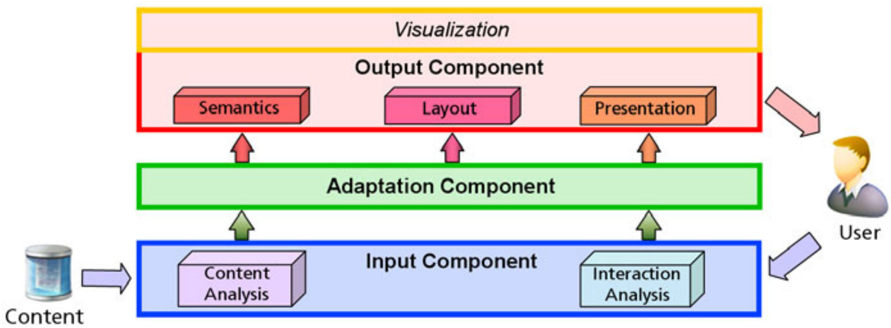


Fig. 1. Reference Model for Adaptive Visualization Systems (MAVS)

The *Input Component* receives different inputs, which are used as implication factors for the *Adaptation Component*. It consists of two different modules, for analyzing these factors: The *Content Analysis Module* prepares the underlying data for the visual presentation and extracts certain criterions and properties of the data for the Adaptation Component. The *Interaction Analysis Module* is responsible for extracting information from different sensors about the user and prepares it for an appropriate usage in the Adaptation Component.

The *Adaptation Component* receives information of the Input Component and adapts several modules of the Output Component accordingly. This includes the selection of the best fitting modules and their parameterization in the Output Component.

The *Output Component* consists of the three modules *Semantics*, *Layout* and *Presentation*. The Semantics Module represents the structure of the data transformed for visualization. Further it contains information about the amount and attributes of the data. The *Layout Module* is responsible for placement and structuring the information on screen (e.g. graph-layout). The *Presentation Component* holds certain visual parameters (e.g. color, size etc.) and provides them to the Layout Module. The instantiation of these three modules together builds visualization for the user.

4 The SemAdapt Framework

In this section, we present *SemAdapt*, a framework for adaptive visualizations based on the concepts described above. This section will focus on the three components input, output, and adaptation.

4.1 Input Component

The Input Component of SemAdapt receives two different impact factors for the visualization adaptation and parameterization: User interactions and data.

Interaction Analysis: The main input component of SemAdapt is responsible for capturing information about the individual user and to extract user information from interaction events. These interaction events appear as natural consequence of the visualization operation and are captured with different contextual information about the interaction type, the layout method and the content. Altogether, an interaction event has the form of a triple `<type, layout, data>`. For example a click with the left mouse button on a graphical representation of the element Fraunhofer produces the following triple

```
(device.mouse.button.left.click, semadapt.semap,
thing.organizations.research.appliedsciences.Fraunhofer)
```

The gathered user interactions serve as input for an instantiation of the Interaction Analysis Module that extracts information about the user behavior. Based on a quantitative analysis on different grades of abstraction, the instantiated module captures the preferences of the user according to the input device, the visualization type and the content [25]. Additionally, the module calculates predictions with the KO*/19-Algorithm and detects user activities as recurring and similar interaction sequences [26]. All this information about the user behavior is passed to the Adaptation Component of SemAdapt to select a visualization type and an individual parameterization of the output modules.

Content Analysis: Beside the user interactions, the data as input and impact factor for the adaptation is considered in SemAdapt. The semantically annotated data can contain geographic or time-dependent information, which is used to identify visualizations that are able to visualize these factors in an appropriate way. Further the amount of the data and its structure is considered. A high amount of ontological concepts forces to choose a visualization that figures the hierarchy in an abstract way and provides a higher-level of interactions with data. Beginning with a search in knowledge domains, the data is analyzed in different steps building a hierarchy and extracting the relevant attributes of data (e.g. timestamps and geographical attributes). The procedure of the data analyzing is described in [27] in detail.

4.2 Adaptation Component

The *adaptation component* is the central component of the interactive system presented in this paper. In order to control the whole system, the adaptation component has interfaces to all components and modules. It gathers information from the Input

Component and transforms it for the Output Component. The gathered user information are persisted in a three dimensional user model [28], that contains tables with preference values for the activities, visualization types and contents.

The whole adaptation process is based on the (static) *visualization capability model* and the (dynamic) *user preferences model*. For each visualization the *visualization capability model* defines its ability to present certain data types (e.g. time-dependent data or concept hierarchy), activities and additional features like time or spatial dependent data. This model is static because the capabilities of a visualization are predefined and do not change.

The adaptation of the Output Component is performed in two steps. In a first step, based on the visualization capability model, the set of visualizations capable to present the user preferences with respect to visualizations, content, and activities are selected. In a second step, these visualizations are ordered according to the visualization preferences of the user. Herewith, the adaptation engine defines the overall layout of the User Interface as well as the parameters for each module in the Output Component. For instance a certain visualization type, the SeMap [22] is chosen in the Layout Module and the chosen visualization is parameterized (e.g. color, placement, size of the icons etc.) with the Presentation Module.

4.3 Output Component

Visualizations can be described according to the following characteristics: *what* is displayed, *where* is it displayed, and *how* is it displayed? Accordingly, each visualization in SemAdapt implements the modules semantics (data), layout and presentation.

The *semantics module* (*data module*) defines which data is visualized. It contains information about the data (what is the data about), its structure (e.g. hierarchy, incoming and outgoing relations) and amount. The main task of this module is to parameterize the data and provide the opportunity to adapt visualization in the lowest level. In semantically annotated data this module has further the separation functionality of different abstraction level (e.g. concept and instance level). With the information about data, data-related adaptation can be realized, e.g. content recommendation or adaptation of level of details.

The *layout module* defines where and how the data will be visualized. Dependent on the user preferences, different graphical metaphors like text, different graph-layout algorithms may be chosen. For each metaphor, different layout algorithms like cone tree or tree-map are provided. It should be noted that only the geometrical layout is defined on this layer, but not yet the visual appearance.

The *presentation module* defines more precisely how the data will be presented. It is the visual layer of the visualizations and parameterizes the visual look, by setting e.g. the texture, color, or size. The input of the presentation is the geometry as calculated by the layout module. The output component of SemAdapt contains a set of different visualizations. Each of these visualizations has its own pipeline of semantics (data), layout and presentation. The visualizations are grouped into three categories:

General visualized components (GVCs) are abstracted visualizations and form the UI. GVCs are primarily responsible for the selection, placement and initialization of visualization types in their layout layer. Their semantics layer separates the data from the visualization. It decides which part of the data should be visualized with which visualization. The presentation layer is responsible for the separation of the visual representations. It just decides which visualizations should have the same visual representation. It is necessary e.g. for comparative visualization. Further the GVCs decide the interactive connections between the visualization. If a user interacts with visualization v1, should the visualization v2 react to the user interaction?

Semantics visualization components (SVCs) visualize the structure of the semantic data and provide the possibility to interact. SVCs are the real visualization components and consist of one or more visualization algorithms, e.g. force directed or concentric radial. While SVCs get their data from the Semantics Layer of the GVCs, their own Semantics layer decides about the number, level-of-detail and ranking of the data to be visualized. The layout of a SVC has different parameterizations, e.g. the centralization of a topic (in concentric radial) or the order of concepts (in SeMap [22]). The presentation layer parameterizes the visual representation, e.g. by choosing colors or sizes for the instances, relations and concepts.

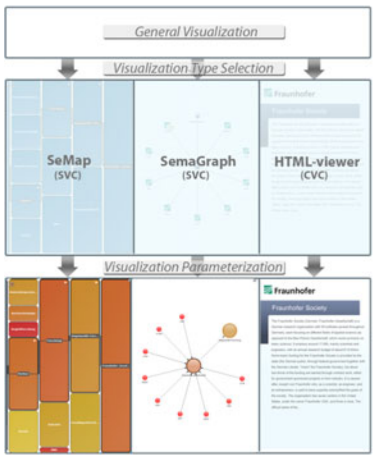


Fig. 1. Three-stepped adaptation of the output component

Content Visualization Components (CVCs) are responsible for the presentation of the content referenced by the semantics. Examples for CVCs are pictures or HTML-viewers. They are interactively coupled to one or more SVCs and inherit the semantics, presentation and layout form theses coupled SVCs. The output component provides a three-stepped adaptation of the visualization. The figure above shows these three steps. In the first step the GVC provides a number of visualization types based on the input and the adaptation component. In the figured case two SVCs (SeMap and SemaGraph) and one CVC (HTML-viewer) are chosen by the GVC. In the next step each of the chosen visualizations creates its own Layout and visualizes the same information in different abstraction levels. Therefore an adaptation of the Semantics (content) and the Layout is performed. The last step adapts the visual representation (Presentation layer) of the information. In the given case, size, order and color are adapted and parameterized to the computed requirements of the user and data respectively.

Several visualizations for SemAdapt were already developed and are used for adaptive visualization of semantically annotated data. These visualizations were introduced in some of our previous works [14,22,23,24].

5 A SemAdapt Application Example

In this chapter we describe the functionality of the SemAdapt framework based on a knowledge exploration example. Knowledge exploration is an important process for adopting knowledge with information systems, whereas graphical representation of the knowledge can help to optimize the learning process and reduce the cognitive overload [14]. For evaluating the system with ground-truth data, we are using the Freebase data base [8] extended with an own schema. Freebase contains open linked data, which can easily be mapped to a formal data specification, whereas *Domains* are the highest concepts (of an ontology), *Types* inherit from *Domains*, the individuals are called *Topics*. The relations are defined as *Properties*. The starting point of our example is the interaction of a user on a graphical representation of the concept *Fraunhofer*, which is analyzed in the Input Component as described above. This interaction produces a search query on the freebase data base. The results are analyzed based on the procedure described in [27].

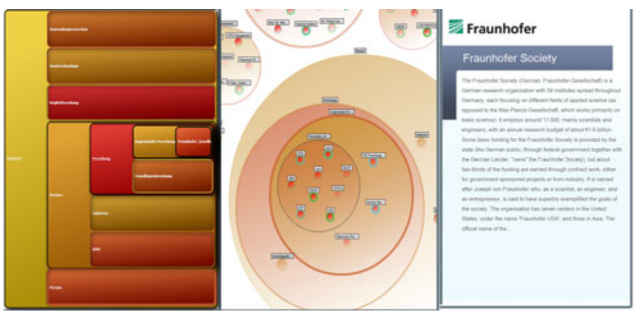


Fig. 4. Example for Visualization Adaptation

The results are information about the data, its structure and the amount of results in different semantics categories. Based on these results a set of visualizations is chosen, that meets the requirements of the analyzed data. Further the interaction is analyzed in a user-centered way based on the procedure described above. The results are a current set of information about the user and the data, which is further compared with the user model described in [28]. The result of the comparison is a table with scalar values for the fitting of different visualization, abstracted data layer (ontological concepts) and activities. Based on the computed values several visualization types and parameterizations are set, whereas a user interface may consist of more than one visualization [14].

6 Conclusion and Future Work

In this paper a reference model and framework for adaptive visualization has been presented with the following main innovative features. The overall conceptual model is based on the distinction of three main components and herewith, supports perspectives to adaptive visualization. In combination with the modular structure of the system's conceptual architecture, this allows the goal-oriented adaptation of specific parts of the system with appropriate level of detail. The user interface consists of a so-called visualization cockpit, which integrates different visualizations, chosen and adapted according to the user preferences. This concept supports both independent views as well as user preferred interactions in closely coupled visualizations. SemAdapt, so far, focuses on the adaptation of visualization type and its parameterization.

One main area of future research is the extension of the framework to handle raw data and metadata. The cockpit concept will be extended so that different data types can be visualized in parallel. Major aspects of future work are editing and annotation and their synchronization along different data types. From an adaptation point of view, this will lead to a substantial extension of both the user model as well as the capability model. Another principal extension of SemAdapt will be the conceptualization of adaptive visualization in Web 2.0, where user groups with different roles and preferences view and edit in shared environments.

References

1. Keim, D.A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H.: Visual Analytics: Scope and Challenges. In: Simoff, S.J., Böhlen, M.H., Mazeika, A. (eds.) Visual Data Mining. LNCS, vol. 4404, pp. 76–90. Springer, Heidelberg (2008)
2. Hui, B., Gustafson, S., Irani, P., Boutilier, C.: The need for an interaction cost model in adaptive interfaces. In: Proceedings of the Working Conference on Advanced Visual interfaces, AVI 2008, pp. 458–461. ACM, New York (2008)
3. Cao, Y., Theune, M., Nijholt, A.: Modality effects on cognitive load and performance in high-load information presentation. In: Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI 2009), pp. 335–344. ACM, New York (2009)
4. Thomas, J.: Visual Analytics: a Grand Challenge in Science-Turning Information Overload into the Opportunity of the Decade. Keynote talk. In: IEEE InfoVis 2005, Minneapolis (2005)
5. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual Analytics: Definition, Process, and Challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008)
6. Brusilovsky, P.: Adaptive navigation support. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 263–290. Springer, Heidelberg (2007)
7. Ahn, J.-w., Brusilovsky, P.: Adaptive visualization of search results: Bringing user models to visual analytics. Information Visualization 8/3, 167–179 (2009)
8. Danine, A., Lefebvre, B., Mayers, A.: TIDES - Using Bayesian Networks for Student Modeling. In: ICALT 2006, pp. 1002–1007. IEEE Computer Society, Los Alamitos (2006)
9. Noguez, J., Sucar, L.E.: A Probabilistic Relational Student Model for Virtual Laboratories. In: ENC 2005: Proceedings of the Sixth Mexican International Conference on Computer Science, pp. 2–9. IEEE Computer Society, Los Alamitos (2005)

10. Anderson, C.R., Domingos, P., Weld, D.S.: Relational Markov Models and their Application to Adaptive Web Navigation. In: KDD 2002: Proceedings of the Eighth ACM SIGKDD 2002, pp. 143–152 (2002)
11. Davison, B., Hirsh, H.: Predicting Sequences of User Actions. In: Predicting the Future: AI Approaches to Time Series Problems, pp. 5–12. AAAI Press, Menlo Park (1998)
12. Leuski, A., Allan, J.: Interactive information retrieval using clustering and spatial proximity. User Modeling and User Adapted Interaction 14(2-3), 259–288 (2004)
13. Gotz, D., Wen, Z.: Behavior-driven visualization recommendation. In: IUI 2009: Proceedings of the 13th International Conference on Intelligent User Interfaces, pp. 315–324. ACM, New York (2009)
14. Nazemi, K., Burkhardt, D., Breyer, M., Stab, S., Fellner, D.W.: Semantic Visualization Cockpit - Adaptable Composition of Semantics-Visualization Techniques for Knowledge-Exploration. In: Proc. 13th International Conference ICL on Interactive Computer Aided Learning, ICL 2010, Hasselt, Belgium (2010)
15. Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization: Using Vision to Think, 1st edn. Morgan Kaufmann, San Francisco (1999)
16. May, T., Kohlhammer, J.: Towards closing the analysis gap: Visual generation of decision supporting schemes from raw data. Computer Graphics Forum, 911–918 (2008)
17. Aaltonen, A., Lehikoinen, J.: Refining visualization reference model for context information. Personal Ubiquitous Comput. 9, 381–394 (2005)
18. Heer, J., Card, S.K., Landay, J.A.: Prefuse: a toolkit for interactive information visualization. In: Proceedings of the SIGCHI. ACM, New York (2005)
19. Brodlie, K.W., Brooke, J., Chen, M., Chisnall, D., Hughes, C., John, N., Jones, M., Riding, M., Roard, N., Turner, M., Wood, J.: Adaptive Infrastructure for Visual Computing. In: Proc. of Theory and Practice of Computer Graphics (2007)
20. Arens, Y., Hovy, E.: The design of a model-based multimedia interaction manager. Artif. Intell. Rev. 9(2-3), 167–188 (1995)
21. Kao, T.-H., Yuan, S.-M.: Designing an XML-based context-aware transformation framework for mobile execution environments using CC/PP and XSLT. Computer Standards & Interfaces 26(5), 377–399 (2004)
22. Nazemi, K., Breyer, M., Hornung, C.: SeMap: A Concept for the Visualization of Semantics as Maps. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 83–91. Springer, Heidelberg (2009)
23. Stab, C., Breyer, M., Nazemi, K., Burkhardt, D., Hofmann, C., Fellner, D.: SemaSun: Visualization of Semantic Knowledge based on an improved Sunburst Visualization Metaphor. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, pp. 911–919 (2010)
24. Stab, C., Nazemi, K., Fellner, D.W.: SemaTime - Timeline Visualization of Time-Dependent Relations and Semantics. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Chung, R., Hammound, R., Hussain, M., Kar-Han, T., Crawfis, R., Thalmann, D., Kao, D., Avila, L. (eds.) ISVC 2010. LNCS, vol. 6455, pp. 514–523. Springer, Heidelberg (2010)
25. Nazemi, K., Stab, C., Fellner, D.W.: Interaction Analysis for Adaptive User Interfaces. In: Huang, D.-S., Zhao, Z., Bevilacqua, V., Figueroa, J.C. (eds.) ICIC 2010. LNCS, vol. 6215, pp. 362–371. Springer, Heidelberg (2010)
26. Nazemi, K., Stab, C., Fellner, D.W.: Interaction Analysis: An Algorithm for Activity and Prediction Recognition. In: Proc. 2nd IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2010. IEEE, Los Alamitos (2010)
27. Nazemi, K., et al.: Interacting with Semantics: A User-Centered Visualization Adaptation based on Semantics Data. In: Jacko, J.A. (ed.) HCII 2011, Part I. LNCS, vol. 6761, pp. 491–500. Springer, Heidelberg (2011)
28. Nazemi, K., et al.: Modeling Users for Adaptive Semantics Visualizations. In: HCII 2011, Part I. LNCS, vol. 6761, pp. 491–500. Springer, Heidelberg (2011)

A Proposal of an Interactive Music Composition System Using Gibbs Sampler

Akira Shirai and Tadahiro Taniguchi

Ritsumeikan University,
312 RacoonD3 3-1-7 Nojihigashi Kusatsushi Shiga,
525-0058, Japan
{shirai,taniguchi}@em.ci.ritsumei.ac.jp

Abstract. In this paper, we propose a novel method for generating a melody from a generative probabilistic model by using Gibbs sampler. Furthermore, users can modify the composed music by interacting with the generation process. This method enables users to create their favorable melodies. Recently, music composition using computer software is gathering attentions. Many people want to compose their original music. However, musical composition is still too difficult for beginners to obtain their favorable original music. Our system highly supports such people easy to create own music. We also evaluate the probabilistic music composition method by several experiments.

Keywords: automatic music composition, interactive system, probabilistic model.

1 Introduction

Recently, a music composition using a computer named DTM is very popular means of the musical activity. People can make an original music easily. However, a music composition is still too difficult for the people who have less knowledge of music. They can't assemble a music structure because of complexity of itself. Therefore, it is a very important to develop the system that assists such a people to create music. Although there are many methods for supporting beginners, we focus on an automatic music composition system.

Automatic composition techniques have been investigated in the past. A selection of keywords is a very popular means for such system. It generates melodies based on keywords (e.g. sad, happy, jazz, rock) that an user select. In this means, the user chooses not feelings of melodies in it but keywords corresponding to feelings. In short, the user converts feelings into keywords and then the system translates these into melodies. However, the user doesn't get needed melodies when the translation is irreversible because of a variety of interpretations. In order to solve this problem, the system using interactive evolutionary computation (IEC) has been proposed [1]. In this system, a melody is expressed as a gene. The user listens to melodies and evaluates them. Then, the system uses a crossover algorithm to generate next generation melodies. After a number of iterations, melodies change into preferable music for the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 490–497, 2011.
© Springer-Verlag Berlin Heidelberg 2011

user. This approach is highly suitable for beginners because of easy process of composition. However, there is a problem that the user is exhausted in evaluation process by reason of a total time of listening generated melodies. On the other hand, a system is proposed that using dynamic programming named Orpheus [2]. This system is designed with considering composition as an optimal-solution search problem under constraints of the prosody of the Japanese lyrics. Users can compose their original music with lyrics automatically and easily on this. However, it is difficult to search for a solution when global constraints are used. Moreover, in the same conditions of former generations, the system cannot generate various types of music due to take an optimal-solution. To solve these problems, we focus on a music generation from statistical models [3]. An example of this approach is IDyOM [5]. It is a model of melody perception based on n-gram model. Moreover, melody generation from this model using Metropolis-Hastings algorithm, a type of Markov Chain Monte Carlo (MCMC) sampling method, is reported. Although this model shows a good performance on the music segmentation, generated melody is not so creative. We think a lack of the code progressions causes this failure. It is one of the most important aspects of the music structure. Another problem is the generation process that is based on entirely the learnt models. In a word, people can't be involved in the generation process. Melodies generated by such a process are almost good, however these are useless when the user is not interested in. The users need their own original and favorable music.

In this paper, we propose a new interactive system for music composition. Our system generates melodies by using Gibbs sampler, a type of MCMC sampling method, from a probabilistic language model according to an n-gram model that learns existing music. Users can interact with the process of generation under constraints given by several musical structures.

We describe the whole system in Section 2. In Section 3, we show some experimental result using our system and conclude in Section 4.

2 Interactive Music Composition System Using Gibbs Sampler

Our system generates melodies using Gibbs Sampler from bigram model with constraints given by grouping structures of a user's preferred song. bigram model learns from music corpora and then grouping structures are calculated by using exGTTM. Furthermore, users can revise the generated melody by pointing a dissatisfaction part of it. Fig. 1 shows a flowchart of this system. Its details are described in sections below.

2.1 Music Corpora

To get probabilities of bigram model, we define music corpora that have notes and code progression, these are divided into eighth note, from existing music. Notes are represented as $\mathbf{s} = (s_1, s_2, ..., s_N)$ each s has a MIDI note number, rest or duration where N is a dimension that expresses length of melodies. Code progression is $\mathbf{c} = (c_1, c_2, ..., c_N)$ with a code name.

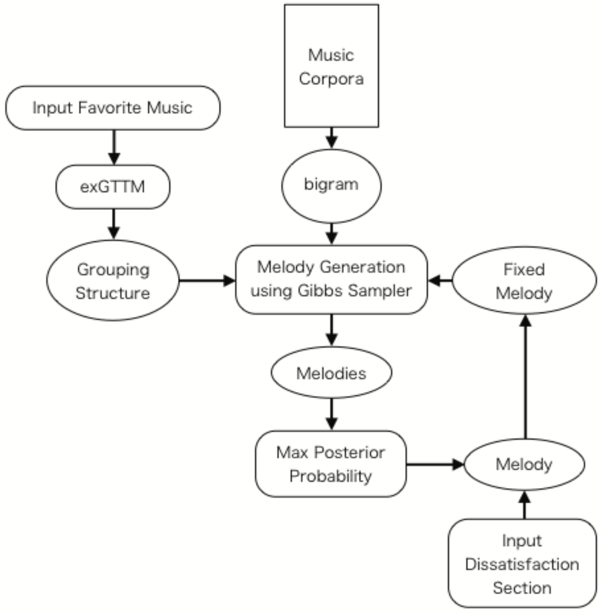


Fig. 1. Flow chart of our system. It generates a melody with favorite music input and music corpora.

2.2 n-Gram Model

We use n-gram model with $n = 2$ (bigram models) as the probabilistic music model β. This model is often used as language models, which approximates the distribution over sentences using the conditional distribution of each word given a context consisting of only the previous $n-1$ words. In this paper, we take notes with codes as words and melodies as sentences. Therefore, a melody's posterior probability is given by

$$p(\mathbf{s} \mid \beta) = \prod_{i=1}^{N} p(s_i \mid s_{i-1}, c_i)\,. \tag{1}$$

To training this model, bigram probabilities are calculated by counting notes in the corpus

$$p(s_t \mid s_{t-1}, c_t) = \frac{\mathrm{C}(c_t, s_{t-1}, s_t)}{\mathrm{C}(c_t, s_{t-1})}\,. \tag{2}$$

$\mathrm{C}(\bullet)$ is a function that counts notes. However, it is possible that probability denotes 0 because of inexistent sequence. Therefore, smoothing method is needed for training n-gram models. In this system, instead of (2), probabilities are trained by using

$$\hat{p}(s_t \mid s_{t-1}, c_t) = \sum_{R} w_R\, p_R\,. \tag{3}$$

$$\sum_R w_R = 1. \tag{4}$$

This smoothing method uses weighted summation. w_R denotes weight of each probabilities and p_R denotes the several different probabilities that are marked by R, where $R = \{(\text{uni}, \text{oncode}), (\text{uni}, \text{offcode}), (\text{bi}, \text{oncode}), (\text{bi}, \text{offcode}), (\text{bi}, \text{degree})\}$. For example, $p_{\text{uni,offcord}}$ is calculated by

$$p_{\text{uni,offcord}}(s_t \mid s_{t-1}, c_t) = \frac{\mathrm{C}(s_t) + B}{\sum_t \mathrm{C}(s_t) + B\lambda}. \tag{5}$$

B denotes a baseline probability and λ denotes a number of the kind of notes. $p_{\text{bi,degree}}$ is peculiar. It is calculated by transporting notes to $\mathrm{C}_{\text{major}}$.

$$p_{\text{bi,degree}}(s_t \mid s_{t-1}, c_t) = \frac{\mathrm{C}(\mathrm{d}(s_t, c_t), \mathrm{d}(s_{t-1}, c_t)) + B}{\mathrm{C}(\mathrm{d}(s_{t-1}, c_t)) + B\lambda}. \tag{6}$$

$\mathrm{d}(\bullet,\bullet)$ is a function that transports notes on codes to $\mathrm{C}_{\text{major}}$.

2.3 Generative Theory of Tonal Music

A generative theory of tonal music (GTTM) [6] is a theory of music that is intended to be a formal description of the musical intuitions of a listener who is experienced in a musical idiom. It is composed of four components. *Grouping structure*, one of these components, expresses a hierarchical segmentation of the piece into motives, phrases, and sections. We suppose that feeling of music is connected with *grouping structure*. According to this hypothesis, users can generate a new favorable melody under constraints given by *grouping structure* of a user's preferred melody.

We use *extended GTTM* [7] to make *grouping structure*. This algorithm is GTTM reorganized for computation. Although *grouping structure* is composed of several rules, our system uses these four rules, GPR (*Grouping Preference Rule*) 2a, 2b, 3a and 3d. These are rules of boundaries between groups. Each rules has a different attribute, slur/rest, attack-point, register and length of notes, which are calculated with their respective parameters. Boundaries are represented as vectors; each value is 0 or 1, of N dimensions. Our system sets these vectors as constraints $\mathbf{z}$ of musical structure.

2.4 Probabilistic Model

We define a probabilistic model of generating grouping structure (Fig. 2) that is expressed by

$$p(\mathbf{z} \mid \mathbf{s}, \beta). \tag{7}$$

In this system, we need a posterior probability of melody $\mathbf{s}$. Thus, an equation (7) is transformed by using Bayes' theorem into

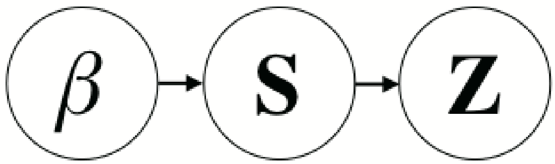


Fig. 2. Graphical model of generation **z**. β is bigram model, **s** is melody and **z** is grouping structure.

$$p(\mathbf{s} \mid \mathbf{z}, \beta) = \frac{p(\mathbf{z} \mid \mathbf{s}) p(\mathbf{s} \mid \beta)}{\sum_{\mathbf{s}'} p(\mathbf{z} \mid \mathbf{s}') p(\mathbf{s}' \mid \beta)} . \tag{8}$$

This can be calculated using equation (1) and conditional probability of **z**

$$p(\mathbf{z} \mid \mathbf{s}) = \frac{1}{(\sqrt{2\pi})^N \sqrt{|\Sigma|}} \exp(-\frac{1}{2} (\mathbf{z} - \mu(\mathbf{s}))^t \Sigma^{-1} (\mathbf{z} - \mu(\mathbf{s}))) . \tag{9}$$

$\mu(\mathbf{s})$ denotes grouping structure of melody **s**.

In order to take grouping structure into the probabilistic model, we suppose that it is generated from *N*-dimensional Gaussian distribution (9).

Since there are λ^N possible melodies, computation is infeasible. However, we can get high probability melody on low complexity by using Gibbs sampler.

2.5 Gibbs Sampler

Gibbs Sampler [4] is a simple and widely applicable MCMC algorithm and can be seen as a special case of the Metropolis-Hastings algorithm. Consider the distribution from which we need to sample. Each step of the Gibbs sampler involves replacing the value of one of the variables by a value drawn from the distribution of that variable conditioned on the values of the remaining variables. The distribution of initial states must also be specified in order to complete the algorithm, although samples drawn after many iterations will effectively become independent of this distribution.

To implement this algorithm into our model, **s** is separated into

$$p(s_i \mid \mathbf{s}_{\setminus i}, \mathbf{z}, \beta) . \tag{10}$$

According to the algorithm, we replace s_i by a value drawn from the distribution $p(s_i \mid \mathbf{s}_{\setminus i}, \mathbf{z}, \beta)$, where s_i denotes the i^{th} note of melody **s**, and $\mathbf{s}_{\setminus i}$ denotes $s_1, ..., s_{i-1}, s_{i+1}, ..., s_N$. Thus, from (1) and (8), we can obtain

$$p(s_i = k \mid \mathbf{s}_{\setminus i}, \mathbf{z}, \beta) \propto p(\mathbf{z} \mid s_i = k, \mathbf{s}_{\setminus i}) p(s_i = k \mid s_{i-1}, c_i) p(s_{i+1} \mid s_i = k, c_{i+1}) . \tag{11}$$

Melodies are generated by using Gibbs sampler algorithm with (11).

2.6 Melody Revision

Automatic composition systems can generate original music. However, it is not creation for users but recommendation from the system. Therefore, composition processes in the system must interact with them.

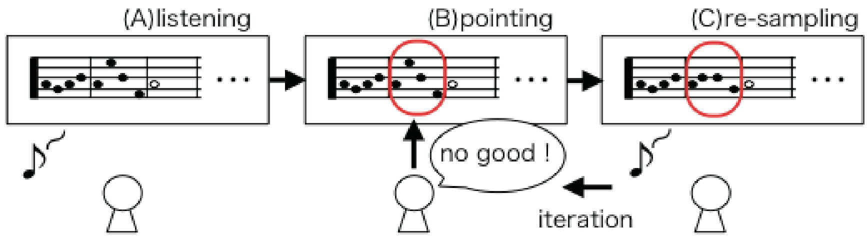


Fig. 3. Revision process of the system. It re-samples part of melody where the user don't satisfy.

In our system, users can revise the generated melody. Revision process (Fig. 3)is executed when the user cannot satisfy with melody. First, the user listens the generated melody and sees the score (Fig. 3 (A)). Next, the user inputs a bar number who feels dissatisfaction (Fig. 3 (B)). Finally, the melody is re-sampled by using Gibbs sampler with cut range of i (Fig. 3 (C)). Revision sampling is affected by $\mathbf{s}_{\setminus i}$ from the property of Gibbs sampler. The user can obtain own original melody after some iteration.

3 Experimental Results

We performed experiments on our interactive music composition system using a 32 verses corpus including 4 Japanese famous musicians. The system generated verse, $N=64$, $B=0.0001$, $T=100$ [1], code progression was Dm7-G-Em-Am and $w_R=\{0.2,0.2,0.2,0.2,0.2\}$.

Fig. 4 shows Mean squared errors of GPRs between generated melodies and user's preferred music. All MSE are reduced on the left in case of $\Sigma=0.05$ compared with the right in case of $\Sigma=0.5$.

Furthermore, we performed evaluation experiments in same conditions except for $w_R=\{0.3,0.1,0.2,0.2,0.2\}$. Subjects selected a preferred music for constraints, and then the system generated melodies. After Gibbs iterations, they listened a generated melody and evaluated it. They could revise it until they felt satisfaction.

Table 1 shows a number of revisions for each subject. Most of subjects revised a generated melody only one time. Surprisingly, subject 4 didn't revise a melody (Fig. 5). It shows that our system can compose interesting melody for users in a short time. However, it is possible that subject 4 hadn't had intention of composition.

---

[1] T: A number of Gibbs iteration.

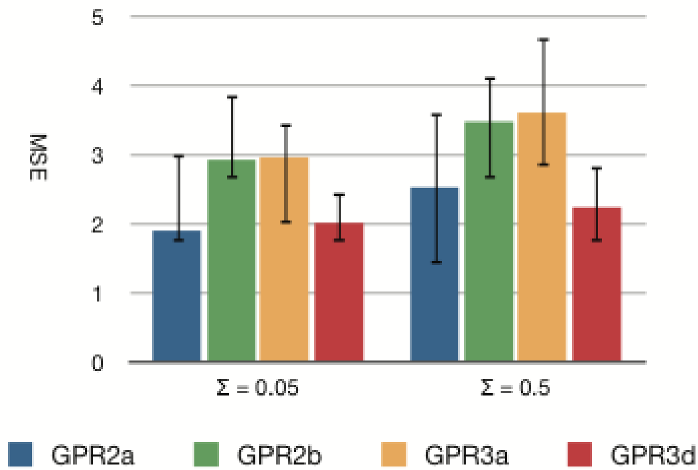


Fig. 4. Mean squared error of GPRs between generated melodies and user's preferred music. Σ is constraint parameter.

Table 1. Results of experiment for subjects. Subject 4 didn't need a revision

Subjects	Artist	Number of revision
Subject 1	Spitz	1
Subject 2	Spitz	1
Subject 3	KinKi Kids	1
Subject 4	**Bump of Chicken**	**0**
Subject 5	Spitz	1

Fig. 5. Generated melody (subject 4). Constraints are given by J-Pop song "Stage of the ground - Bump of Chicken".

4 Conclusion

We have described an interactive music composition system using Gibbs sampler. In addition, we have shown some results of experiments using this system. The experiments showed good results. Generated melodies have less non-harmonic notes. Moreover, the structure of generated melodies is similar to the structure that is given as constraints. The creativity of melodies is another important aspect. It is highly subjective, melodies generated by our system satisfy a user's demand. Most of users that use this system prefer generated melodies. The interaction part also receives a good evaluation. However, this part is claimed with its UI and function. Our system now doesn't have good UI. The user has to input by keyboard on console screen. It is a critical problem for intuitive system that we aim at. The users probably feel troublesome. Thus, a special UI that properly affects melody (when manipulations link with user's intuitions for music, they may not aware of musical structures) is needed.

Music has various aspects (such as pitch, metric, code progression, hierarchical structure and so on) that affect the user's intuition. Therefore, our system must include more and more structural constraints. We develop our system with results described above. The system that associates with the singing voice synthesis system is considerable in the future work.

References

1. Tokui, N.: Music composition with interactive evolutionary computation. In: Third International Conference on Generative Art, Milan, Italy, pp. 215–222 (2000)
2. Fukayama, S., Nakatsuma, K., Sako, S., Yonebayashi, Y., Kim, T.H., Qin, S.W., Nakano, T., Nishimoto, T., Sagayama, S.: Orpheus: Automatic composition system considering prosody of Japanese lyrics. In: Natkin, S., Dupire, J. (eds.) ICEC 2009. LNCS, vol. 5709, pp. 309–310. Springer, Heidelberg (2009)
3. Conklin, D.: Music Generation from Statistical Models. In: Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences, Aberystwyth, Wales, pp. 30–35 (2003)
4. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelliegence 6(1), 721–741 (1984)
5. Wiggins, G.A., Pearce, M.T., Mullensiefen, D.: Computational modeling of music cognition and musical creativity. In: Dean, R. (ed.) Oxford Handbook of Computer Music and Digital Sound Culture. Oxford University Press, Oxford (2009) (in press)
6. Lerdahl, F., Jackendoff, R.: A generative theory of tonal music. MIT Press, Cambridge (1983)
7. Hamanaka, M., Hirata, K., Tojo, S.: Implementing "A Generative Theory of Tonal Music". Journal of New Music Research 35(4), 249–277 (2006)

Sensing User Needs: Recognition Technologies and User Models for Adaptive User Interfaces

Barnabas Takacs[1,2], Lajos Simon[1], and Matthias Peissner[3]

[1] Semmelweis Medical University,
1094 Budapest, Ferenc tér 15, Hungary
[2] Technical University of Budapest,
1111 Budapest, Magyar Tudosok Krt 2., Hungary
[3] Fraunhofer IAO, Nobelstrasse 12,
70569 Stuttgart, Germany
Btakacs@digitalCustom.com, simon@psych.sote.hu

Abstract. In this paper we introduce the notion of a Virtual User Lab that employs virtual reality tools to simulate End-Users in realistic application scenarios in order to help industrial designers and application developers to create and test adaptive interfaces that evolve as users' preferences and potential handicaps are discovered. We describe key elements of the VUL, discuss computer vision-based algorithms for facial information processing to understand user behavior and present an email-reading scenario to better highlight the system's adaptive capabilities and practical usability.

Keywords: User interface adaptation, facial information processing, virtual human interface.

1 Introduction

One of the key challenges in the process of developing adaptive user interfaces is not we not only need to sense the users' interaction and emerging patterns but further derive information about their intent and possibly internal state as an estimation of their likes and dislikes about the content being shown. This requires consistent interpretation of multiple sensory signals that are available in a given situation and interpret these signals in the context of the content being shown along with personalized user profiles. During the past decades, researchers have tried many input modalities to derive such information, but many of these failed due to the lack of a unified modeling and interpretation framework that has helped designers to test their products against life-like, but simulated situations.

To overcome this problem, we have designed a concept, called the *Virtual User Lab* (VUL) that integrates the toolset of virtual reality and humanoid animation with an abstraction layer for various external sensor modalities, all delivering pieces of information about the behavior of a user while accessing the utilities provided by any specific application. This paper describes the architecture and sensory capabilities of the VUL and details how we process this information to help drive the underlying user interface adaptation processes, which is the main focus of our research.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 498–506, 2011.
© Springer-Verlag Berlin Heidelberg 2011

The classical treatment of *Human Computer Interaction* (HCI) considers the process of interaction between man and machine as somewhat *unequal* opponents [1]. The human element, i.e. the user, is involved in two basic processes: *Perception* and *Control*, while the computer's perspective on the other hand is frequently considered not as much along the lines of cognitive and intentional behavior, but as simple, low-level processes implemented in the form of *agents*. To make this pair equal the same modalities of communication readily available for people shall be implement for computers as well in order to give them a full set of senses complete with goals and cognition. This basic model of human computer interaction first assumes is that there are minimally two separate ***agents*** involved, one human and one machine. They are physically separated, but are able to exchange information through a number of information channels. On the side of the human user there are two basic processes involved, namely ***Perception*** and ***Control***. The ***Perceptive process*** may be further divided into a) Huma*n Input Channels* (HIC) and b) *Computer Output Media* (COM) while similarly the ***Control process***, comprises of c) *Human Output Channels* (HOC) and d) *Computer Input Modalities* (CIM). For a multi-modal *HCI* system that possesses an adaptive user interface that constantly improves during time, are especially interesting and challenging. Our goal is to create an intelligent user interface which is able to interpret the user's actions, decide for itself the best way to present information to the user, and cope with one of several input modalities selected by the user without the need to explicitly tell the computer.

User behaviors (like interest, boredom, excitement, confusion, fatigue, etc.) are accompanied by different patterns of facial expression, eye-gaze, head nod, hand movement, gestures and body posture. These so called *surface level behaviors* can be interpreted as indicators of "*on-task"* and *"off-task"* [3] in a general terminology and could be used as measures that help drive interface adaptation processes. On a greater scale, body postures convey specific meanings regarding the actions of the person in front of the output device (a TV or a computer screen), while the *direction of eye gaze* and *head orientation*, for example, are prime indicators of a person's focus of attention. More evidently *spontaneous facial expressions* and head nods are also good indicators of motivational and affective states in relation to the application. In particular, approving head nods, facial actions like smiles, tightening of the eyelids while concentrating, eyes widening, or raising eyebrows suggest interest and excitement (*on-task*). On the other hand, head shakes, the lowering of eyebrows, nose wrinkling and depressed lower lip corners indicate *off-task* behavior.

Clearly, these visually described behaviors often are easier to verify using other types, of specialized, non-visual sensors that are also more acceptable for End-Users protecting their privacy. Either case the output of these sensors must be interpreted in the context of an application itself and the system must form a consistent representation of the user's intention as it changes over time. Based on these observations we argue that virtual reality tools in combination with multi-modal sensory interfaces and advanced visual processing may be used to simulate and present such conditions and thus help industrial designers and application developers to better test how their designs will behave in realistic situations. This idea, what we call the Virtual User Lab, is the topic of the following section.

2 The Virtual User Lab

The goal of the *Virtual User Lab* (*VUL*) is to create a virtual environment supporting industrial users in the development, simulation and evaluation of adaptive user interfaces for user groups with certain needs. Major features of the *VUL* focus on specific areas to *i)* to illustrate user profiles and their changes over time, *ii)* to visualize possible user interface adaptations, *iii)* to simulate specific user interface designs and adaptation processes as experienced by a user with specific limitations, and *iv)* to create a pattern browser in order to provide designers and developers with an overview of design guidelines and an easy access to application specific design patterns which serve as reusable components for adaptive user interfaces. It is a multi-purpose key module of a more complex user interface system, which on one hand, serves as a *virtual reality environment* to access and visualize adaptation processes that take place in various *application* scenarios, and in another sense contains a *virtual user* that may be used to simulate user actions in response to a specific interface and verify adaptation processes during the design of such products or services. In order to keep its maximum flexibility the *VUL* was also designed to interface to external programs and sensors that may be used to measure user reactions in-vivo and perform adaptations on the fly to prove the strength of different adaptation strategies. Finally, the same virtual user also serves as a tool to show designers how a person with impairments or limited physical capabilities would end up using their design and how adaptation processes would take place under those conditions. This is demonstrated in Fig 1.

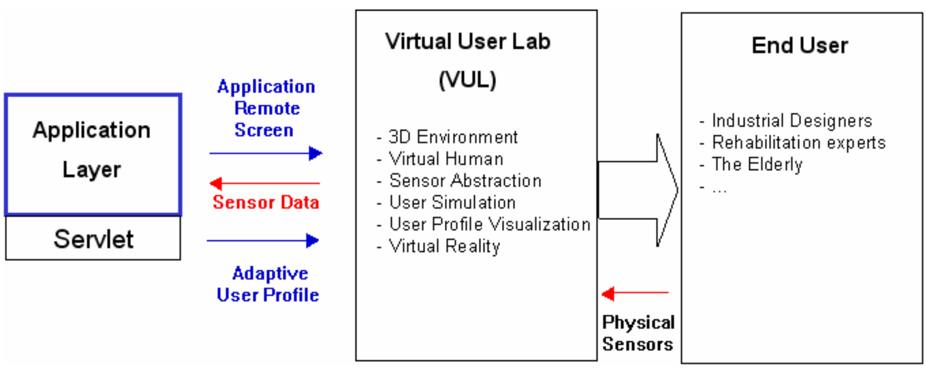


Fig. 1. Generic architecture and operation of the Virtual User Lab

2.1 Physical Sensors vs. Data Abstraction Layer

As in real-life application variety of external sensors may be used to indicate user presence, actions and intentions, which all provide similar types of information, but all differ in their physical manifestation and device characteristics, we first integrated an abstraction layer for an array of sensor types that generally correspond to capabilities of human senses, such as vision, hearing and tactile or motion-based feedback. The raw *input data is* processed by specialized algorithms is categorized into four distinct groups. We call this the Virtual Human Interface as demonstrated in Figure 2.

The most important of these data streams are the images received from one or more video cameras. These are processed by a *computer vision* module, which primarily looks for faces, facial features, body gestures or other telltale signs of the user's presence, his or her actions or internal state. These algorithms correspond to a real web camera mounted on the top of the TV or alternatively viewing a virtual user in its own 3D environment and producing simulated virtual camera views for the adaptation algorithms to process or intelligent cameras and visual devices that provide already processed information on 3D and segmented components, such as the *Kinect*. The second set of algorithms employs *signal processing and audio analysis* to process data captured by the microphones. *Tactile and Motion interfaces* are implemented via touch screen and acceleration-based motion sensors, such as the *Wii*. Finally *physical devices* with no direct perceptual meaning are simply mapped further as raw data. Irrespective what channel the information is arriving from, the purpose of processing is to "recognize" and in some aspect "understand" the its meaning (e.g. to verify the presence of a face in front of the TV and analyze user's motion to arrive at inferences for adaptation purposes). The output of these mechanisms therefore can be mapped onto attributes of virtual *markers* that carry this information from the physical domain into an abstract symbolic representation suitable for the low-level reactive scripts and a high level *Adaptation Engine* (not described in this paper) to operate on.

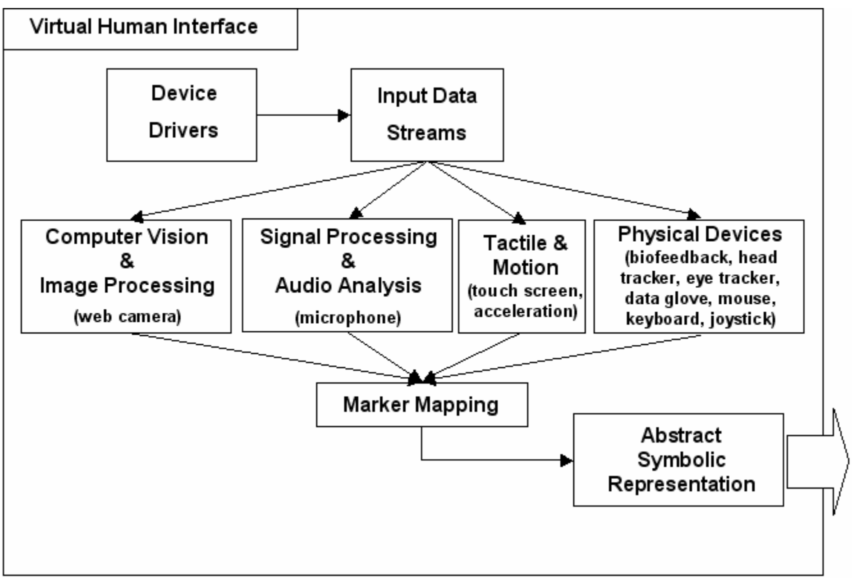


Fig. 2. Functional diagram of the Virtual Human Interface module responsible for processing sensory information

The module described above provides a real-time link between the end-user and the *VUL* via a *symbolic* representation of the outside world. This symbolic representation encodes events as they occur in the physical environment of the user and as an attempt to "understand" what is happening to him or her in the course of the interaction. Irrespective, however, of what the initial source of the data is, the essence of meaningful information it supports is encoded in the form of *markers*. Markers are dimensionless data holders with physical location and attributes in the synthetic 3D environment of the virtual human in the *VUL*. They carry unique *hierarchical labels* and *attributes values* encoding physical locations or abstract properties and they can be referred to by their names when describing high-level tasks in a rule-based behavior control system. They are used as input to the set of rules that control the adaptation algorithms of the entire framework.

To demonstrate how such abstraction is used to derive information about the user, in the next section we will discuss in more detail the visual processing capabilities of the VUL. Clearly these results maybe generalized to other classes of sensors as well.

3 Facial Information Processing

Most facial processing systems attempt to recognize user behavior, most particularly emotions on a person's face as a one-shot process using high level holistic models of how faces look or low-level feature extraction mechanisms that map directly onto emotions using neural networks or other types of statistical classification algorithms [3,4]. In the *VUL* instead we created an algorithm that is based on forming a hypothesis with the help of low level cues and gradually refines that hypothesis using contextual and behavior rules. As a simple example, one may consider a simple *face detector* that is used the estimate the size of the head and in a correlated fashion the distance of the user's head from the screen. Such information could not only be useful for a HCI system to be able to adapt, e.g. the font size of an email reading application, but also provides clues whether the user can see what is on the screen well enough or not. Specifically, it has been suggested that a *lean forward gesture* may indicate the lack of clear visibility and the system could change the rendering of the messages in response.

Taking this idea further we can also analyze the facial regions as determined by the face detector and look for tell-tale signs of emotions, attention, etc. *Region-based image processing* is a basic technique identifying temporal events as they occur on the face. They can be executed in fast and more importantly in parallel to offer the possibility of collecting events that may (or may not!) signify a certain facial actions. Thus the *VUL* system views these events as "indicators" and resolves their potential meaning with the help of the *emotion constraints* known from psychology [5]. As facial tissue moves under the control of the muscles, darkened skin-lines in the form of edges frequently appear perpendicular to the direction of motion. These secondary facial features, such as sharply occurring temporary ridges due to the folding of skin, may be localized and their effect quantified. These features, called *furrows and wrinkles*, vary with each person, yet their very presence is strongly associated with the meaning of facial expressions. As an example, furrows in the upper facial area may be indicate that the user is thinking or trying to resolve/understand something very hard so we need to bring up a Help window. By using a face detector as mentioned above to locate the face and a rudimentary geometric model of facial regions, such as fore head, nose, mouth, etc., one may use localized image processing algorithms to measure the changes in these regions. In particular, these furrows and wrinkles are relatively easy to detect using real-time computer vision methods as they from rather characteristic *horizontal* and *vertical edges*. In our solution the image captured by the web-camera is first divided into facial regions using the eyes and nose detected and the sub-images corresponding to these facial regions of the upper-, middle and lower face, respectively are normalized before being passed on to the edge detection unit. The edge detection algorithm employs standard image processing filter operators and convolution masks in combination with morphological operations.

The operations carried out by these *low-level image processing* algorithms are demonstrated in Figure 3. Here the normalized images, the output of the binary operators, and the respective processing convolution kernels are shown. Different kernels are used to detect horizontal, vertical or diagonal lines. The outputs of these kernels are then combined into a single interpretation to provide a global measure describing the degree of furrows appearing in that portion of the image. On the left high gradient components of a face during a neutral and expressive state, respectively in the forehead and eye regions are shown, while on the right the output of the same operators for the mouth, chin and cheek regions are displayed. Besides *edge detection*, other real-time operators, such as *optic flow* and *morphological* filtering may also yield measures of widened eyes, a smile or open jaw. Figure 4 shows the typical output of these signals at the marker levels. While these signals individually are inherently noisy, together they provide a reliable measure that gradually enables the system to learn certain facial expressions and subsequently use them to better understand the users' needs.

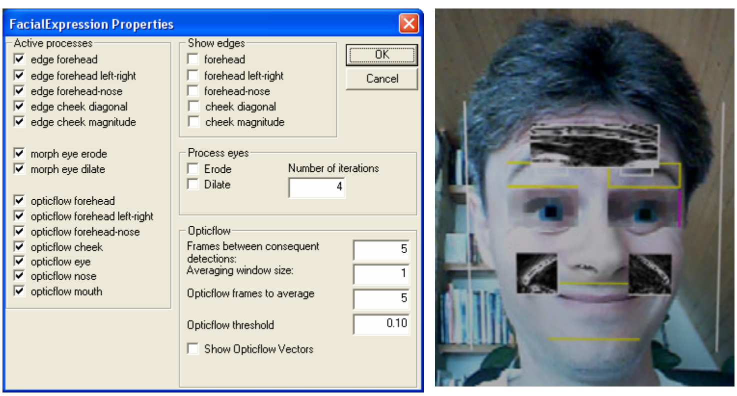


Fig. 3. Facial expression analysis using low-level image processing operators in specific regions of the face

To interpret these competing individual systems we map them on a circular, psychologically validated representation consistent with how humans express emotions during natural interaction. The set of competing rules attributed to each marker action shift the initial estimate of a neutral face in the direction of a specific emotion with a likelihood associated with the intensity of the facial event observed. The outcome of this process is a single estimate of the user's face indicating the most likely emotion he or she is in.

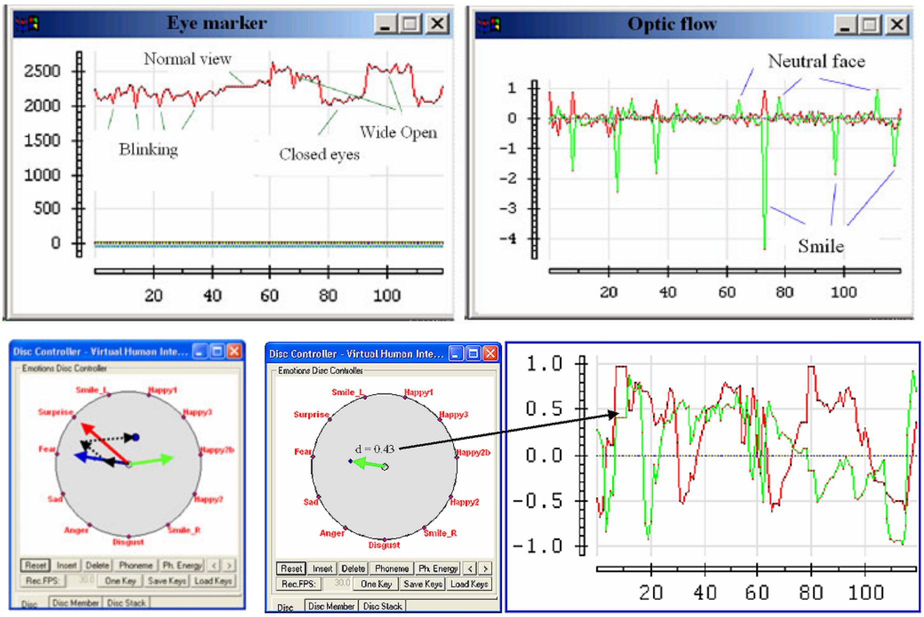


Fig. 4. Markers derived from consistent visual cues indicate the likelihood of local events on the face and in turn allow for a global interpretation of facial signals

4 Practical Application Scenario

The practical use of the Virtual User Lab hereby is demonstrated in a scenario designed for email reading applications using a remote adaptive server. In Figure 5 the main components of our system are shown. The GMail account the user is connected to renders the information contained in the email messages adaptively based on the user's behavior pattern. To achieve this goal the *VUL* captures the remote email screen and displays it within one of two virtual environment, on a TV set in the virtual home and on the projector screen at the research lab, in this case. In both scenarios, the 3D environment allows the VUL to simulate different lighting and visibility conditions, as well as effect, like glare or even visual impairments of the *End-User* viewing this application. The user sees the email messages on a computer screen, or an industrial designer may experience it in a virtual reality headset (VR) for better immersion. The virtual environment is fully panoramic 360 view with many interactive elements most importantly a *3D virtual remote controller*, that may be used to control the remote application by pressing buttons. This virtual remote also allows the application to output physical infra red (IR) control signals and thus change channels on a physical device, as an example.

The Physical sensors attached to the *VUL* in this scenario are a web camera responsible for processing facial information, an ultrasonic distance sensor (as a second alternative to provide calibrated user distances and recognize gestures, such as leaning forward) an RFID reader to help identify different users and set preferences accordingly and finally a physical TV remote the signals of which are captured by the VUL and transmitted to the remote servlet simultaneously for adaptation purposes. In parallel to the physical signals, the facial information processor detects the face and outputs its estimated position and distance from the screen used again, to detect body gestures that may indicate poor visibility on the user's account.

The system is capable of recording and playing back all markers for testing, simulation and validation purposes. Designers interested in how their physical devices perform (e.g. the TV remote) under such conditions may create various visibility tests, realistically simulate visual impairments that rely on 3D distance information and rendering parameters and of course application adaptation rules on the server side may also be tested via simulated user interaction.

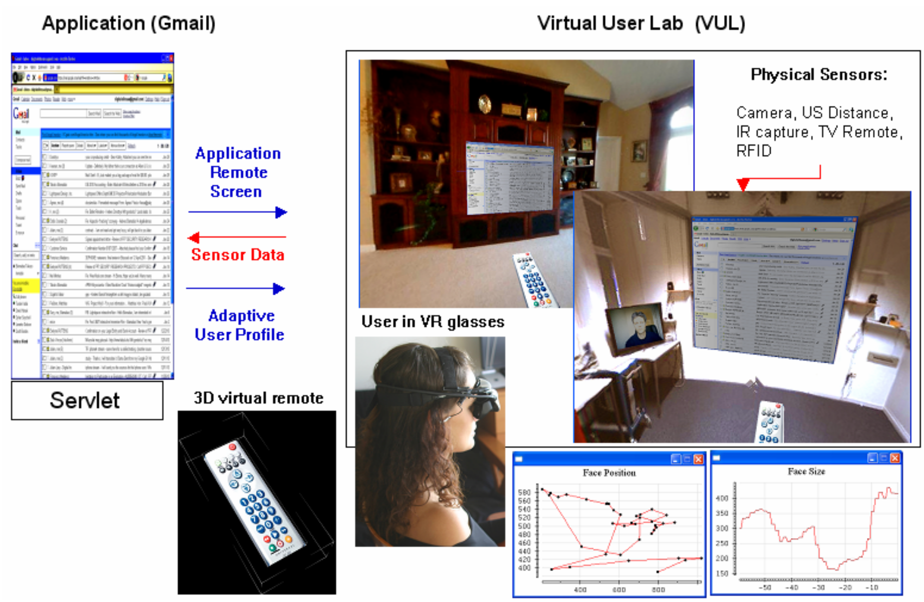


Fig. 5. Application scenario of the virtual user lab in a remote email reading case study using a NetTV and a virtual remote controller (see text)

5 Summary

In this paper we presented a novel approach to help user interface designers to create and test adaptive interfaces using a simulation tool, called the Virtual User Lab. Our approach combines the benefits of virtual reality to render a remote application in a virtual scene, change its viewing conditions and rendering quality both in terms of lighting and visibility factors or handicap as well as user behavior. The latter is measured with the help of a variety of local sensors that may be as simple as a web cam or more complex, like distance and motion sensors or even biofeedback. We have interfaced our solution with a client-server application model and demonstrated its use in a remote Email reading application.

References

1. Schoemaker, L., et al.: Multimodal Integration for Advanced Multimedia Interfaces. In: ESPRIT III, Basic Research Project 8579 (2004)
2. Kapoor, A., Mota, S., Picard, R.W.: Towards a Learning Companion that Recognizes Affect. In: Proc. of Emotional and Intelligent Agents II: The Tangled Knot of Social Cognition, AAAI Fall Symposium (2001)
3. Bruce, V.: Face Recognition. In: Eyesenck, M.W. (ed.) Cognitive Psychology: An International Review (1992)
4. Kort, B., Reilly, R.: Analytical Models of Emotions, Learning and Relationships: Towards and Affect-sensitive Cognitive Machine. In: Proc. Intl. Conf. on Virtual Worlds and Simulation, San Antonio, TX (2002)
5. Schlossberg, H.: The Description of Facial Expressions in Terms of Two Dimensions. J. of Experimental Psychology 44(4) (1952)

uPlatform: A Customizable Multi-user Windowing System for Interactive Tabletop

Chenjun Wu, Yue Suo, Chun Yu, Yuanchun Shi, and Yongqiang Qin

Department of Computer Science and Techonology, Tsinghua University, Beijing, P.R. China
anwingwu@gmail.com, suoyue@tsinghua.edu.cn, yc2pcg@gmail.com, shiyc@tsinghua.edu.cn, qyq08@mails.tsinghua.edu.cn

Abstract. Interactive tabletop has shown great potential in facilitating face-to-face collaboration in recent years. Yet, in spite of much promising research, one important area that remains largely unexplored is the windowing system on tabletop, which can enable users to work with multiple independent or collaborative applications simultaneously. As a consequence, investigation of many scenarios such as conferencing and planning has been rather limited. To address this limitation, we present uPlatform, a multi-user windowing system specifically created for interactive tabletop. It is built based on three components: 1) an input manager for processing concurrent multi-modal inputs; 2) a window manager for controlling multi-user policies; 3) a hierarchical structure for organizing multi-task windows. All three components allow to be customized through a simple, flexible API. Based on uPlatform, three systems, uMeeting, uHome and uDining are implemented, which demonstrate its efficiency in building multi-user windowing systems on interactive tabletop.

Keywords: tabletop, multi-user, windowing system, window management.

1 Introduction

In recent years, tabletop system has shown great potential in facilitating face-to-face collaboration. Yet, in spite of much promising research, one important area that remains largely unexplored is the windowing system on tabletop. Familiar in desktops, windowing system enables user to work with several programs simultaneously. It is also necessary for tabletops because many scenarios such as conferencing, planning, and exhibition require parallel interaction on multiple independent or collaborative applications. This requirement is being further magnified with the emergence of large-scale interactive tabletops [1, 2] where users may have more chance to work independently.

However, there arise several challenges when replanting the windowing system from desktop to tabletop due to three fundamental factors of tabletop: multi-user usage, direct input and horizontal display. First, the co-located multi-user support of tabletops brings many new problems, such as authorization and conflict handling. These issues are inherently ignored by current windowing systems such as Microsoft

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 507–516, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Windows and X Window, because they are designed for single user usage. Second, the direct input makes the WIMP interfaces of these most popular windowing systems unsuitable for tabletops since multi-touch interfaces lack the precision usually afforded by indirect pointing devices. Third, the horizontal display requires more research to be performed to understand space management practices on tabletop.

Little research has been performed to investigate these challenges in a systematic way. While a number of improvements for current windowing systems have been proposed by HCI researchers in order to explore space management [3], or support multi-point input [4, 5], very few carefully consider the issues raised in co-located multi-user usage, such as authorization and conflict management. On the other hand, nor has work been performed to help researchers investigate these challenges. Most of current toolkits for tabletops [2, 6, 7] only focus on full-screen, single application development. Researchers must spend excessive effort on developing the underlying plumbing before they can focus on their interested areas, such as input authorization.

To address this limitation, we present uPlatform, a multi-user windowing system specifically created for interactive tabletop. Our driving goal was that this system would be customizable enough to facilitate the design of innovative user interfaces and window management policies in the context of multi-modal input, multi-task UI and multi-user usage, and at the same time be simple enough for average researchers to quickly learn and use. Therefore, uPlatform is a tool for creating new types of tabletop environment rather than a new tabletop proposal. It is built based on three components: 1) an input manager for processing concurrent multi-modal inputs; 2) a window manager for controlling multi-user policies; 3) a hierarchical structure for organizing multi-task windows.

This paper is organized as follows. After introducing some related work, we describe uPlatform by providing an overview of its architecture as well as a description of its API and configuration tools. We then illustrate its customizability through three sample systems. Finally, we conclude with a discussion on our experiences and findings.

2 Related Work

Many works focus on extending the functionality of current windowing systems in order to support groupware applications. The MIDDesktop [4] provides a desktop-like environment to execute several Java applets simultaneously. The Swing Applets simply execute in sub-windows within the MIDDesktop display area, and can receive inputs from multiple mice. MPX [5] is a modified X Window System that can accommodate multiple input devices, and provides new APIs for developers to create collaborative applications based on these inputs. SDGToolkit [8] brings multi-mice support to Microsoft Windows operating system. It also considers the orientation issues for tabletop displays and can automatically rotate the cursor according to a participant's seating angle. However, the shortcoming of these systems is that their paradigm is too closely tied to desktop environment and mice input, which makes their WIMP interfaces unsuitable for tabletops since multi-touch interfaces lack the precision usually afforded by these indirect pointing devices.

There are also several tools targeting towards helping HCI researchers explore the novel window management policies. Ametista [9] is a mini-toolkit that supports the creation of new OpenGL-based desktop environments using both a low-fidelity approach using placeholders, as well as a high-fidelity approach based on X app redirection. Similarly, the Mettise [3] toolkit uses an image compositing approach that makes it possible to apply a number of visual effects and geometrical transformations on windows. As one of its examples, Mettise shows how it enables tabletop interface that features automatic window orientation. Unfortunately, both Ametista and Mettise do not take input customization into consider and only support single user interaction with traditional window applications using mouse.

Various Toolkits [2, 6, 7] have been built specifically for the interactive tabletop. They all provide abstractions to support the core tabletop interaction functionality, allowing researchers to concentrate on their applications. For example, the DiamondSpin toolkit [6] implements "around the table" interaction techniques to address orientation challenges in tabletop use. The T3 toolkit [2] allows multiple tabletops to be connected together to support mixed-presence collaboration. PyMT [7] provides a growing collection of reusable widgets and interaction techniques in order to accelerate the development of applications on tabletops.

Unfortunately, all these toolkits assume that the tabletop system can only display a single application at one time, and the API they provide is based on the notion of a single rectangle display with no layout information. As the result, it's impossible to use these toolkits to set up scenarios such as meeting and brainstorming that need to execute multiple collaborative applications simultaneously, unless we modify the source code or create an ad-hoc layer that host and manage other applications. However, both the approaches are a hard task, one that few HCI researchers are willing to do.

3 The uPlatform System

We design uPlatform for several reasons: Most importantly, we were unable to find a tabletop environment that supports executing multiple independent or collaborative applications simultaneously on interactive tabletop. It is a crucial requirement for our current research, since we are exploring many novel scenarios such as meeting and planning on a large-scale tabletop of size 2m*4m. These scenarios all involve mixed-focus collaboration [10], where individuals frequently transition between individual and shared tasks within a group. Therefore, we wanted a multi-user windowing system specifically created for interactive tabletop. We also definitely wanted the system to be customizable enough so that HCI researchers can explore innovative user interfaces and window management policies in the context of multi-modal input, multi-task UI and multi-user usage. Furthermore, as a tool for creating new types of tabletop environment, uPlatform should be efficient enough, allowing for rapid prototyping and experimentation.

In this section, we provide an overview of the uPlatform architecture, followed by a discussion of the implementation, a description of the customizability of each component, and the development and configuration tools provided by uPlatform.

3.1 The Architecture

uPlatform's architecture was designed with two main themes in mind: simplicity and extensibility. Figure 1 depicts a schematic of the architecture and execution of the main system loop. uPlatform is built based on three components: 1) an input manager

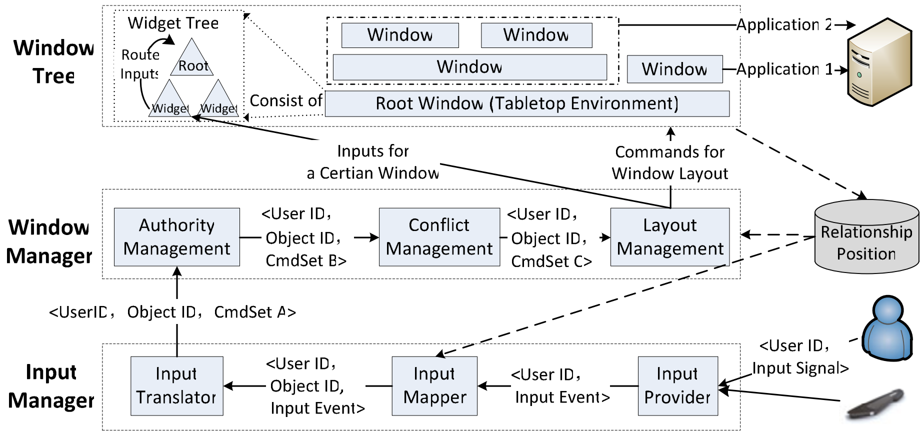


Fig. 1. The architecture of uPlatform

for processing concurrent multi-modal inputs; 2) a window manager for controlling multi-user policies; 3) a hierarchical structure for organizing multi-task windows.

Input Manager. The input processing is much more complicated on tabletops than that on desktops. Typically, multiple users provide input streams from different channels, including hands, styluses or physical objects. In addition, the same input signal can be translated in different ways. For example, different gesture libraries can be defined for different tasks. Therefore, it is important to separate higher level application logics from multimodal inputs to leave developers or researchers focusing on the design of interface components.

In uPlatform, we divide the input processing procedure into three steps: 1) input providing step to receive and normalize input events from different input devices; 2) input mapping step to determine which window an input action intends to operate on; 3) input translating step to encode raw inputs signals into primitive commands before they are sent to window manager for further processing. Each step can consist of one or several objects to do the concrete work. For example, the input providing step may contain TUIO provider and multi-mice provider at the same time.

Window Manager. The main task of Window Manager is to coordinate the inputs from different users, and change the position and relationship of windows based on the inputs and context. According to this, we assign the tasks of windows management to three sub components: authority manager, conflict manager and layout manager, which deal with command authorization, conflict resolution, and windows layout respectively. Input commands will pass through these three managers in order and each manager has the ability to stop the inputs from further traversing.

In order to obtain flexible and extensible management, the managers all work in a policy-based way. Formally, policy is an abstract representation of the properties and behaviors that define how the management is done. Each policy provides sufficient interfaces that can be used and extended by researchers. In uPlatform, we define three types of policies to deal with authority, conflict and layout issues respectively. Each window can hold these three policy instances to define how its child or offspring windows are managed.

Window Tree. Like most windowing systems on desktop, uPlatform uses a tree structure to organize all the windows into a hierarchy. Each application based on uPlatform consists of one or several windows. The root of this tree is a full-screen application, which is called *Tabletop* and should be designated before the underlying windowing system starts to run. Other applications can then be loaded and hosted within the root window. Each window is encapsulated into an independent thread, which holds its own message loop and rendering engine. Therefore, windows can respond concurrently to multiple users' simultaneous input.

Also like many GUI toolkits, each window is made up of a widget tree, which can be seen as a WPF control, and created visually in Microsoft Visual Studio or other design tools. A number of operations are supported by uPlatform to manipulate windows. For example, rotate, reposition a window, maximize, restore a window or change the parent of a window.

3.2 Discussion of Implementation

We implement uPlatform based on WPF and C# programming language. WPF is Microsoft's next generation graphical subsystem on Microsoft Windows platform that enables modern UI features such as transparency and transforms. These features, together with powerful IDE provided by Microsoft, give great convenience for users building rich, rotatable applications. Besides, WPF provides a good framework for reusing, customizing and extending existed components, which is crucial for an extensible development tool.

We implement a WPF control called *UObject* as the root of each window. It holds the basic look and feel of the window, such as border and menu. To achieve multi-layer window organization, we define a Canvas control at the top of UObject which can hold child windows. In addition, each UObject is located within a *HostVisual* that holds its own event dispatcher and works independently with other windows. We monitor each dispatcher carefully, and catch any exception thrown out of its processing loop in order to guarantee the whole system running correctly.

Since WPF's input processing mechanism does not work in a multi-thread environment, we implement the input manger from the ground up. All input events inherit from UTable.Inputs.InputEventArgs class with fields such as device ID and user ID. After the inputs are mapped and sent to a window with the help of input mapper and layout manager, uPlatform is responsible for routing the inputs from a certain leaf widget to the root widget so that each widget along the path has the opportunity to handle the inputs and stop further routing. This is similar to WPF's input routing mechanism for handling inputs within a window.

3.3 uPlatform API

uPlatform provides a simple, flexible API for customizing the input manager, window manager and user interfaces discussed above. This API is packed into a development toolkit called *uTableSDK* [11], which is freely available to academic researchers. In this section, we describe a few of the most important classes.

Customize Inputs. Inputs can be customized by implementing classes that inherit from *IInputProvider*, *IInputMapper* or *IInputTranslator* interface, and putting them into uPlatform's input manager. These three interfaces define how a certain type of input is provided, mapped and translated respectively. Take IInputProvider as example, the only method it contains is the InputTriggered event whose type is listed below:

```
delegate void InputEventHandler(InputEventArgs args);
```

The customized input provider can raise this event with any input argument that inherits from InputEventArgs at any time, which will be captured by input manager, and sent to the input mappers. The InputEventArgs in fact has no concrete fields, and it depends on researchers to add data such as multi-touch position or key type. Researchers can add or remove providers simply by calling the following two methods in UTableHelper class:

```
RegisterInputProvider(Type inputType, IInputProvider p)
UnregisterInputProvider(Type inputTy, IInputProvider p)
```

Customize Window Management Policies. *LayoutPolicy*, *ConflictResolvePolicy* and *AuthenticationPolicy* are base classes for researchers to modify the management policies of window manager. They each contain several template methods that will be called by uPlatform when commands are passed through the corresponding managers. For example, LayoutPolicy provides:

```
void OnInputReceived(IObject obj, InputEventArgs args)
void OnObjectInserted(IObject obj)
```

These two methods will be called to handle layout inputs on a certain child window or define how to put an incoming child window respectively. Therefore, the researchers have the freedom to define how the windows should be rearranged according to user input.

Unlike interfaces in input manager, the policy classes can be set in each window. It means different windows can hold different management policies, which gives even more flexibility. Management policies can be changed dynamically by simply calling the following methods in the UObject class:

```
LayoutPolicyType = typeof(DemoLayoutPolicy);
LayoutPolicyParameter = null;
```

Customize. User Interfaces. uPlatform allows researchers to rapidly create tabletop content in a way similar to implementing a WPF window: edit the widget tree in a WYSIWYG manner in Visual Studio. In addition, we provide a rich set of multi-touch controls and Visual Studio templates so that the development can be further accelerated by simple drag-drop under Visual Studio.

3.4 Development and Configuration Tools

uPlatform integrates well into Visual Studio, and provides an efficient and comfortable development environment to researchers. For example, researchers can simply open a project template in Visual Studio in order to create a new uPlatform-powered system.

uTableSDK provides an xml file for each system to configure environment settings (Fig 2). Because most of our development occurs on desktops, the system by default is configured to use multiple mice as its input provider, and use a blank window with physical layout policy as the root tabletop. These settings can be changed easily when we want to test the application in different environments. For example, change the input provider to TUIO when this application is deployed on an interactive tabletop.

```
<UTable>
    <!--Define the root of window tree and its management policies-->
    <Tabletop>
     <Class Type="UTable.Objects.UHome" Assembly="UTable.Objects"/>
     <LayoutPolicy Type="UTable.Policies.PhysicalPolicy" Assembly="UTable.Objects">
       <Property Name="LinearDamping" Value="0.6"/>
     </LayoutPolicy>
     <ConflictPolicy Type="UTable.Policies.RankPolicy" Assembly="UTable.Objects"/>
     <AuthorityPolicy Type="UTable.Policies.PublicPolicy" Assembly="UTable.Objects"/>
    </Tabletop>
    <!--Define the inputs-->
    <Input><Providers><Provider>
    <Class Type="UTable.Input.MultiMouseProvider" Assembly="UTable.Input"/>
  </Provider><Providers></Input>
</UTable>
```

Fig. 2. Customize windowing system through a config file

4 Examples

In this section, we will discuss three windowing systems constructed based on uPlatform. They demonstrate the variety of tabletop user interfaces and management policies that uPlatform can facilitate. All systems are developed within one day, which demonstrate the framework's efficiency in building multi-user multi-task windowing systems on horizontal displays.

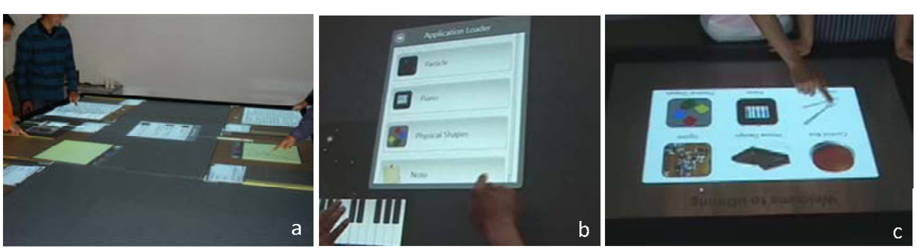


Fig. 3. The windowing systems based on uPlatform. (a) uMeeting, (b) uHome, (c) uDining

4.1 uMeeting

The first windowing system is called uMeeting (Fig.3a). It is developed for meeting scenario, where multiple users work individually or collaboratively on a large-scale interactive meeting table. This system shows the advantage of the flexible UI management using the layout policies, authority and conflict policies of uPlatform.

We create a UObject to represent the workspace that is located directly on the tabletop. The workspaces divide the tabletop into multiple sub-regions, and work as a working unit for each user. User-defined applications, such as document, photo and note can be put on the workspace, and transferred between tabletop and different workspaces. uMeeting implements a physical policy that is applied to the tabletop and workspaces with different parameters so that windows can be dragged and thrown with friction and inertia. Personal authority policy is applied to workspace to ensure that only the owner has the authority to manipulate the content of his personal space. The tabletop, however, is assigned public authority policy so that every user has the identical authority of the objects directly put on tabletop.

4.2 uHome

The scenarios for interactive tabletop at home, such as tea table, are quite different from those in a meeting room. Firstly, the display scale is smaller, and participants can see and interact with any part of the table; secondly, the activities on the tabletop usually involve collaboration of multiple users based on one or several applications; finally, the activities may change frequently, especially when there are some guests. Noticing these, we create a system, called uHome, to help manage applications on an interactive table at home (Fig.3b). It uses a simple two-layer structure to organize the user interface: the bottom is tabletop and the upper layer contains applications running on the system. A special application, called AppLauncher can be launched from the pie menu of the tabletop, which can be used to start other applications in this system. uHome replants uMeeting's public authority policy and the first-come, first-use conflict resolution policy to the tabletop of uHome so that everyone have the equal opportunity to manipulate the applications.

4.3 uDining

uDining is a windowing system custom made for an interactive dining table. When designing uDining, we focus on some specific characteristics a dining table has. For example, the participants sit at the four sides of the table, and involve in a single task, such as ordering food. Based on these observations, we apply a simplest layout policy to uDining: a single application is full-screen displayed on the table. All inputs are sent to the active application except for the special buttons at the center of the table, which can trigger a global menu. One main command in this menu is switching the active application to AppBrowser, a native application of uDining, as shown in Fig.3c. This application gives users a way to browser and selects all applications installed on the system.

Since users in uDining only interact with applications, not the management system itself. We do not assign any authority and conflict control policy to uDining. In fact, we believe in these scenarios, it's the application's task to care about these issues.

5 Discussion

The three systems described in the last section illustrate the basic capabilities and extensibility of uPlatform. Through our experience developing windowing systems for different interactive tabletops based on uPlatform, we noticed many issues related to policies design and replantation among different scenarios and devices. The following are a few of them that belong to part of our current investigation.

Although the screen scale and scenarios vary greatly for different tabletops, their windowing systems still share a lot of features on inputs, management policies and user interfaces. uPlatform takes advantage of this and greatly reduce the cost when implementing different windowing systems. In fact, the native input utilities we provide in uPlatform can satisfy most of the input requirements. Furthermore, since we can reuse the management policies and widgets in previous systems, the difficulties and time decrease significantly when our work go deeper.

The extensibility also plays an important role in implementing different windowing systems. We have an impressive experience when we want to add a pen as an input device to the system. The typical way is to add a multi-touch provider to the system so that application can respond similarly to other touch input without changing anything. However, we can also create a provider with a new input channel so that applications can respond differently.

The policies also give a good way for extending previous work. In practice, there are two ways to do this. One is reusing a previous policy with different arguments. For example, the tabletop and workspace can both have physical property while the fiction on them can be different. The other way is inheriting from a policy and rewrite or extend its logic. This way can be more powerful with a little more effort.

6 Conclusion

In this paper, we have described uPlatform, a customizable multi-user windowing system specifically created for interactive tabletop. We have presented the general architecture of the system and described its API and configuration tools. We have presented several examples of uses that demonstrate its efficiency in building multi-user windowing systems on interactive tabletop.

The SDK of uPlatform is available from http://utablesdk.codeplex.com.

Acknowledgment. This paper was supported by Development Plan of China under Grant No. 2009AA01Z336 and National Natural Science Foundation of China under Grant No. 61003005.

References

1. Jens, T., Marc, H., Benjamin, W., Lasse, S., Sebastian, F., Markus, K., Rainer, M.: Advancing Large Interactive Surfaces for Use in the Real World. In: Advances in Human-Computer Interaction, vol. 2010, pp. 1–11 (2010)

2. 1hilip, T., Peter, R.: T3: Rapid Proto-typing of High-Resolution and Mixed-Presence Tabletop Applications. In: Proc. TABLETOP 2007, pp. 11–18 (2007)
3. Chapuis, O., Roussel, N.: Metisse is not a 3D desktop! In: Proc. UIST 2005, pp. 13–22. ACM Press, New York (2005)
4. Shoemaker, G.B.D., Inkpen, K.M.: MIDDesktop: An Application Framework for Single Display Groupware Investigations. School of Computing Science, Simon Fraser University, Report No. TR 20001-01 (2001)
5. Hutterer, P., Thomas, B.H.: Enabling Co-located Ad-hoc Collaboration on Shared Displays. In: Proc. AUIC 2008, vol. 76, pp. 43–50. Australian Computer Society Inc. (2008)
6. Shen, C.: DiamondSpin: An Extensible Toolkit for Around-the-Table Interaction. In: Proc. CHI 2004, pp. 167–174. ACM Press, New York (2004)
7. Thomas, E.H., Juan, P.H.: PyMT: A Post-WIMP Multi-Touch User Interface Toolkit. In: Proc. TABLETOP 2009, pp. 17–24. ACM Press, New York (2009)
8. Tse, E., Greenberg, S.: Rapidly Prototyping Single Display Groupware through the SDGToolkit. In: Proc. AUIC 2004, vol. 28, pp. 101–110. Australian Computer Society Inc. (2004)
9. Nicolas, R.: Ametista: a mini-toolkit for exploring new window management techniques. In: Proc. CLIHC 2003, pp. 117–124. ACM Press, New York (2003)
10. Tang, A., Tory, M., Po, B., Neumann, P., Carpendale, S.T.: Collaborative coupling over tabletop displays. In: Proc. CHI 2006, pp. 1181–1190. ACM Press, New York (2006)
11. uTableSDK, `http://utablesdk.codeplex.com`

Synchronization and Fluctuation of Rhythm in Musical Cooperative Performance

Tomohito Yamamoto

College of Information Science and Human Communication,
Kanazawa Institute of Technology
7-1 Oogigaoka, Nonoichi, Ishikawa, 921-8501 Japan
tyama@neptune.kanazawa-it.ac.jp

Abstract. A live musical performance gives us better impression than recorded music heard from a portable music player. From player's point of view, live performance also gives better impression than playing music with metronome or recorded music. Thus, the difference exists between the live musical performance and the recorded performance that doesn't change in real time. In this research, to clarify the difference, the cooperative performance of the drum and the bass was analyzed from a rhythmical aspect. The results showed that synchronization error between musical rhythms, and fluctuation of musical rhythm became smaller in cooperated performance than in the performance with recorded music.

Keywords: Music, Rhythm, Synchronization, Fluctuation, Cooperative Performance.

1 Introduction

A Portable music player such as an iPod has been widespread recently and people can enjoy music anytime and anyplace. Nevertheless, live musical performance still gives us better impression than recorded music. From player's point of view, live performance also gives better impression than playing music with metronome or recorded sound. These empirical facts suggest that there is qualitative difference between recorded music which never changes at real-time and live musical performance. To clarify this difference contributes not only to develop an artificial musical playing agent but also to understand fundamental human communication mechanism which is quite complicated. In this research, from such point of view, we analyze a cooperative musical performance.

A music player changes musical tempo consciously or unconsciously to give artistic expression to their performance. There are some researches which deal with such temporal change in piano performance [1]-[3]. In those researches, the relation between fluctuation of rhythm and grouping of phrase or artistic expression has been discussed. Recently, not only piano performance but also drum performance has been analyzed. For example, Okuhira et. al. have analyzed sense of groove quantitatively

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 517–526, 2011.
© Springer-Verlag Berlin Heidelberg 2011

which possibly determine whole music expression [4]. Watanabe et. al. have analyzed the relation between standard deviation of drum beat and subjective metronome [5].

While these researches have analyzed change of musical tempo of solo performance, our research group has analyzed a musical cooperative performance. In this research, we analyzed the relation between musical aspect and physiological aspect of piano ensemble, and resulted that synchronization between musical rhythm and respiratory rhythm changed depending on difficulty of performance [6]. Werner et. al. have also analyzed the synchronization of musical rhythm of cooperative piano performance and resulted that proper sound feedback is important to synchronize musical rhythm [7]. In addition to these researches, there are some researches which have analyzed non-verbal information in cooperative performance. Katahira et. al. have analyzed the relation between drum performance and visual information. The results showed that visual information contributed to synchronize musical rhythm [8].

Moreover, with the development of information technology, analysis of interaction between a human player and an artificial player has appeared. Horiuchi et. al. have analyzed cooperative performance of solo player and accompanist using multiple regression analysis. From results, they have developed the control model of musical performance [9]. Kobayashi et. al. have analyzed cooperative performance of two piano players, and resulted that time difference of sounding, and change of time difference affected tempo control. From these results, Kobayashi have developed an artificial ensemble system [10].

A Summary of these previous researches is that in researches about solo performance, fluctuation of musical rhythm has analyzed, and the relation between artistic expression and fluctuation has been discussed. However fluctuation of cooperative performance has not been discussed. On the other hand, in researches about cooperative performance, synchronization of musical rhythm and its mechanism, or non-verbal information and physiological aspect have been analyzed. However the difference between performance with recorded music and cooperative performance has never been discussed.

Therefore, in this research, the difference between performance with recorded music and cooperative performance are analyzed form rhythmical aspect. Especially, the cooperative performance of the drum and the bass which is very popular with jazz, rock and pops and needs rhythmical cooperativeness for good performance are focused.

2 Methods

2.1 Task and Subjects

2 kinds of music were prepared for the experiment (Fig.1(a)-(d)). One was simple music, and another was complex music (BPM =120, 16 bars). For the experiment of simple music, 8 subjects (4 armature drummers and bassists who had 3.6 years playing experience in average.) participated. Similarly, for the experiment of complex music, another 8 subjects (who had 3.3 years playing experience in average.)

participated. In both experiments, 2 bassists and 2 drummers were paired (Therefore 2 pairs were prepared in each experiment.). In each pair, there were 4 combinations. Therefore 8 combinations of a drummer and a bassist were prepared in each experiment. All subjects were instructed to have a practice with a score and recorded CD a few days before experiment. Moreover, just before experiment, subjects had a practice to accustom themselves to the environment and playing the music. Therefore, in this experiment, the technical difference between subjects and environmental effects were regarded to be negligible.

In this experiment, 4 conditions were prepared and 5 performances were conducted in each condition.

Condition 1: Performance with click sound (BPM =120) from a speaker (Click Condition)
Condition 2: Performance with the music which was recorded in Click condition (Rec Condition)
Condition 3: Cooperative performance without visual information which was realized to stand screen between players (Live1 Condition)
Condition 4: Cooperative performance with visual information (Live2 Condition)

Experiment was conducted from 1 to 4 condition. Performance order affected experiment results were also regarded to be negligible because subjects had a practice enough before experiment. Between condition, subjects had enough rest. Therefore fatigue did not affect the experiment results.


(a) Music for the Drum (Simple)


(b) Music for the Bass (Simple)


(c) Music for the Drum (Complex)


(d) Music for the Bass (Complex)

Fig. 1. Music for Experiment

2.2 Experiment System

Fig.3 shows the experiment system. Musical performance was recorded by “Cubase4” (Steinberg). The silent MIDI drum “DTXPROLER” (YAMAHA) and the subjects’ own bass were used for experiment. Performed data of the drum (MIDI data) and the bass (raw wave data) were captured through digital mixer “01X” (YAMAHA). In Rec condition, to play back recorded drum performance, “MOTIF RACK” (YAMAHA) and a Speaker (GX-D90, ONKYO) was used.

In this experiment system, latency (the time delay in each recording device) might occur. Therefore latency of each device was measured. As a result, under 1 msec of latency was observed. This result indicated that latency of this system was negligible.

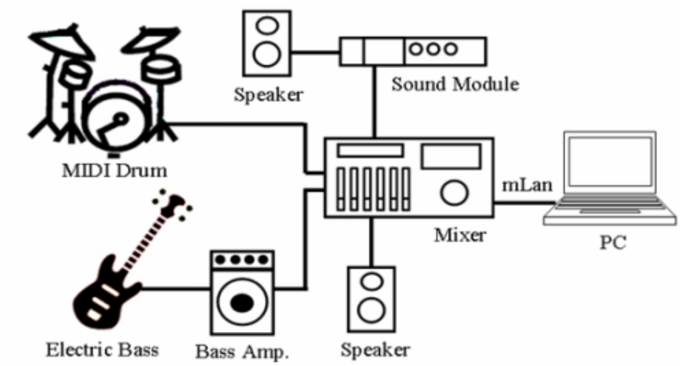


Fig. 2. Experiment System

2.3 Data Analysis

In this research, to clarify the difference between performance with recorded music and cooperative performance, rhythmical aspect of musical performance was analyzed. Concretely, fluctuation of periodical rhythm and synchronization error (S.E.) between 2 players (or between a player and click or recorded music) was analyzed. In addition to this analysis, subjective evaluation to musical performance was conducted by questionnaire.

S.E. was calculated by the method below; at first, the sounded time (or Click sound in condition 1) of the drum was calculated from MIDI data. And the sounded time (or Click sound in condition 1) of the bass was calculated from wave data (In each sound, the first peak of wave form was selected as sounded time.). Next, as shown in Fig.3(a), absolute value of the time difference between corresponding sounded times was calculated as S.E..

In this research, high hat cymbal, snare drum and bass drum were used for performance. However in cooperative performance, a bassist tends to play with the sound of snare and bass drum. Therefore these two drums and corresponding bass sound were used for data analysis (In analysis of Click condition, similar analysis process was adopted.).

Fluctuation of periodical rhythm was calculated by the method shown in Fig.3(b). The formula (2) was for calculation of standard deviation of period, and it was used for index of fluctuation.

For subjective evaluation, questionnaire was conducted after experiment. Question items were “Easiness of performance”, “Enjoyment of performance”, “Subjective synchronization error”. After experiment the drummers and bassists answered these items by 5 grades, and next, they commented about each experiment condition freely.

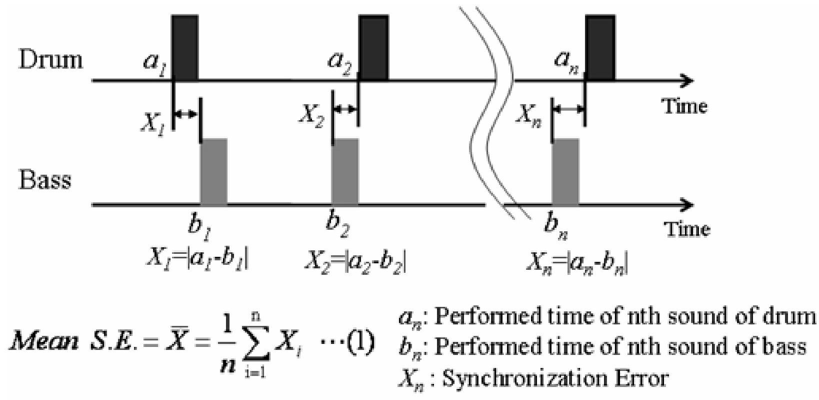


Fig. 3(a). Calculation of synchronization error

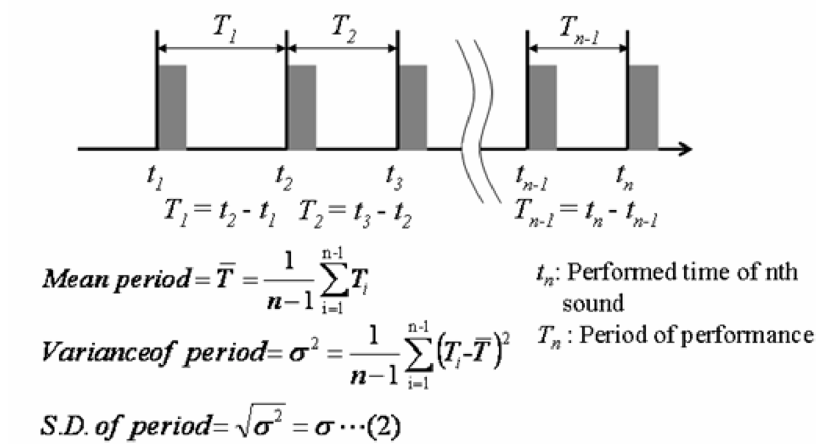


Fig. 3(b). Calculation of fluctuation of periodical rhythm

3 Results

3.1 Results of Simple Music

Fig.4(a) shows mean and S.D. of fluctuation of periodical rhythm which is calculated from data of all 8 pairs (5 performances * 8 pairs = 40 data). In this research, at first, Levene's test was conducted to analyzed data. As a result, the equality of variance was not confirmed. Therefore, including not only data of subsection 3.1 but also data of 3.2, significant difference between medians is tested by Kruskal-Wallis test.

The result of drum performance shows that there is no significant difference between conditions (X^2 (3) = 1.746, n.s.). This result means that drum's fluctuation of periodical rhythm is mostly same in all conditions.

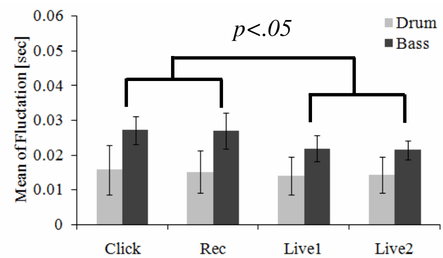
The result of bass performance shows that there is significant difference between conditions (X^2 (3) = 51.4, $p<.05$). Therefore, multiple comparison (Steel-Dwass test) is conducted between conditions. As a result, there is significant difference between Click and Live1, Click and Live2, Rec and Live1, Rec and Live2 (t = 5.102, 5.881, 4.278, 4.783, df = 78, $p<.05$). This result means that fluctuation of periodical rhythm in Live1, 2 condition is smaller than fluctuation in Click and Rec condition in which musical rhythm do not change interactively.

Fig.4(b) shows mean and S.D. of S.E. which is calculated from data of all 8 pairs. The results of Kruskal-Wallis test is that there is significant difference between conditions in drum and bass performance (X^2 (3) = 55.202, 34.339, $p<.05$). Therefore multiple comparison is conducted between conditions. The result of drum performance is that there is significant difference between Click and Live1, Click and Live2, Rec and Live1, Rec and Live2 (t = 3.494, 3.696, 6.376, 6.554, df = 78, $p<.05$).

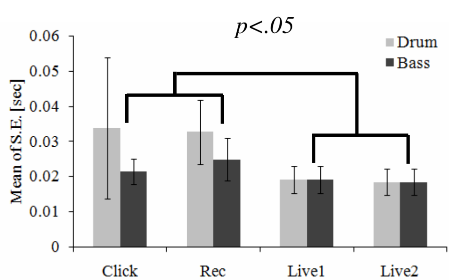
The result of bass performance is that there is significant difference between Click and Rec, Click and Live1, Click and Live2, Rec and Live1, Rec and Live2 (t = -2.637, 2.695, 3.504, 4.437, 4.749, df =78, $p<.05$). The value of Live2 is smallest in all conditions. These results show that S.E. of cooperative performance is smaller than that of Click and Rec condition in which players play according to static musical tempo.

Fig.5 shows the result of subjective evaluation. It is easy to understand that Live2 gets the highest point. The results of Kruskal-Wallis test is that there is significant difference between conditions in all question items ($p<.05$). The result of multiple comparison is that in "Easiness of performance", there is significant difference between Click and Rec, Click and Live1, Click and Live2, Rec and Live1, Rec and Live2 ($p<.05$). In "Enjoyment of performance", there is significant difference between Click and Live1, Click and Live2, Rec and Live1, Rec and Live2 ($p<.05$). In "Subjective S.E.", there is significant difference between Click and Rec, Click and Live2, Rec and Live1, Rec and Live2 ().

These results suggest that cooperative performance give better subjective evaluation, and visual information contribute to get higher value.



(a)



(b)

Fig. 4. Mean and S.D. of (a) fluctuation of period, (b) synchronization error

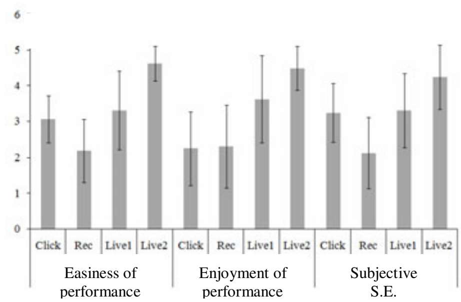


Fig. 5. Mean and S.D. of Subjective evaluation

3.2 Results of Complex Music

Fig.6(a) shows mean and S.D. of fluctuation of periodical rhythm which is calculated from data of all 8 pairs (5 performances * 8 pairs = 40 data). The result of drum performance shows that there is significant difference between conditions (X^2 (3) = 10.786, $p<.05$). The result of multiple comparison is that there is significant difference between Rec and Live2 (t = 3.438, df = 78, $p<.05$). This result means that fluctuation of periodical rhythm in Live2 condition is smaller than fluctuation in Rec condition in which musical rhythm does not change interactively.

The result of bass performance shows that there is no significant difference between conditions (X^2 (3) = 1.501, n.s.). This result means that bass's fluctuation of periodical rhythm is mostly same in all conditions.

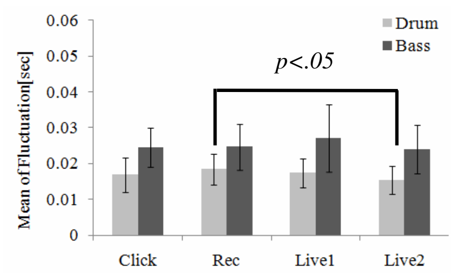
Fig.6(b) shows mean and S.D. of S.E. which is calculated from data of all 8 pairs. The results of Kruskal-Wallis test is that there is significant difference between conditions in drum and bass performance (X^2 (3) = 29.493, 45.997, $p<.05$). Therefore multiple comparison is conducted between conditions. The result of drum performance is that there is significant difference between Click and Rec, Rec and Live1, Rec and Live2 (t = -4.378, 4.888, 3.666, df = 78, $p<.05$).

The result of bass performance is that there is significant difference between Click and Rec, Rec and Live1, Rec and Live2 (t = -6.274, 5.494, 4.496, df =78, $p<.05$). These results shows that S.E. of cooperative performance is smaller than that of Rec condition in which players play according to static musical tempo.

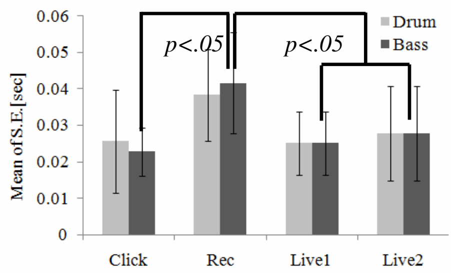
Fig.7 shows the result of subjective evaluation. From results, Live2 gets the highest point in terms of “Easiness of performance” and “Enjoyment of performance”. However in terms of “Subjective S.E.”, Live1 gets the highest point.

The result of Kruskal-Wallis test is that there is significant difference between conditions in all question items ($p<.05$). The result of multiple comparison is that in “Easiness of performance”, there is significant difference between Click and Live2, Rec and Live1, Rec and Live2 ($p<.05$). In “Enjoyment of performance”, there is significant difference between Click and Live1, Click and Live2, Rec and Live2, Live1 and Live2 ($p<.05$). In “Subjective S.E.”, there is no significant difference between conditions.

These results suggest that cooperative performance give better subjective evaluation, and visual information contribute to get higher value in easiness and enjoyment



(a)



(b)

Fig. 6. Mean and S.D. of (a) fluctuation of period, (b) synchronization error

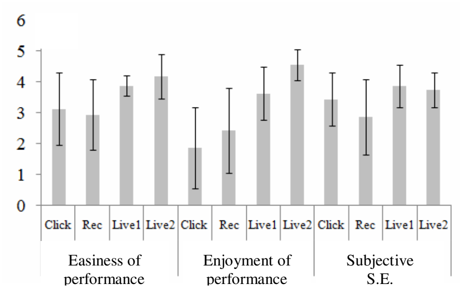


Fig. 7. Mean and S.D. of Subjective evaluation

of performance. However, playing complex music, visual information not always contributes to get higher value.

3.3 Comparison between Simple and Complex Music

The results of simple music and the results of complex music come from different subjects. However, for just reference, these two results are compared in this section.

Comparing Fig.4(a) and Fig.6(a), fluctuation of complex music is bigger than that of simple music without bass performance in Click and Rec condition. The result of Mann Whitney U test between simple and complex music shows that there is significant difference between Rec conditions, and Live1 conditions in drum performance (U = 489, 511, $p<.05$). In base performance, there is significant difference between Click, Rec, Live1, Live2 conditions (U = 568, 578, 525, 567, $p<.05$).

Comparing Fig.4(b) and Fig.6(b), S.E. of complex music is bigger than that of simple music without drum performance in Click condition. The result of Mann Whitney U test between simple and complex music shows that there is significant difference between Click, Live1, Live2 conditions in drum performance (U = 582, 448, 438, $p<.05$). In base performance, there is significant difference between Rec, Live1, Live2 conditions (U = 199, 448, 438, $p<.05$).

4 Discussion

In this research, to clarify the difference between performance with recorded music and cooperative performance, the cooperative performance of the drum and the bass were analyzed form rhythmical aspect. The summary of results is below;

- In most case, fluctuation of periodical rhythm in Live condition is smaller than fluctuation in Click or Rec condition.
- S.E. of Live condition is smaller than that of other conditions.
- Live condition get highest point in subjective evaluation.
- In most case, fluctuation of periodical rhythm in simple music is smaller than fluctuation in complex music.
- In most case, S.E. of simple music is smaller than that of complex music.

The results of S.E. show that two players most synchronize in cooperative performance. If just synchronizing mutual sounds, it is assumed that to synchronize a sound to repetitive click sounds is easiest than other conditions. However cooperative performance whose musical tempo may change gives best results. The reason why S.E. of cooperative performance get good result is that human cannot keep perfect constant rhythm, and as a result fluctuation should be appeared. In cooperative performance, both players predict mutually musical tempo listening partner's fluctuated performance. In this situation, if prediction of players matches, their musical tempo will more synchronize than in playing with repetitive click. In playing with recorded music, mutual prediction never happens, and recorded music has bigger fixed fluctuation than click. As a result, less synchronization is observed.

The results of fluctuation show that cooperative performance which do not have a pace maker give better result than performance with Click. This result is very interesting. This kind of results demands the control mechanism which realizes interaction between players, and at the same time, realizes stability of musical tempo. One possible mechanism is entrainment of nonlinear oscillators. However it cannot explain the result of Click condition which is enforcement entrainment and should give best result in S.E. and fluctuation. In future, it is necessary to develop a dynamical system which explains this phenomenon for modeling.

In the result of subjective evaluation, Live2 condition gets highest point in terms of "Easiness of performance", "Enjoyment of performance". However, in terms of "Subjective S.E.", Live1 gets highest point. These results correspond to the results of S.E., and means that in complex music, visual information do not contribute synchronization error. Moreover these results mean that in Live2 condition, players enjoy their performance, although synchronization error become bigger. The reason of this result is that in Live2 condition, visual information might enhance tension or sense of coexistence. As a result, it affected subjective evaluation of performance.

The result of this research shows that synchronization and fluctuation of cooperative performance is smaller than that of performance with fixed tempo, and at the same time cooperative performance gives better impression than performance with fixed tempo. Described above, it depends on success of mutual prediction of players. To predict timing of sounding is realized by referring past information about performance. This referring ability is for not only musical performance but also general human activity. For example, in conversation, human predict subsequent utterance by past utterances and speak. In future, we would like to analyze the relation between general communication and musical performance.

5 Conclusion

In this research, to clarify the difference between performance with recorded music and cooperative performance, the cooperative performance of the drum and the bass are analyzed form rhythmical aspect. As a result, (a) in most case, fluctuation of periodical rhythm in cooperative performance is smaller than fluctuation in performance with recorded music, (b) S.E. of cooperative performance is smaller than that of performance with recorded music, (c) cooperative performance get highest point in subjective evaluation.

In future works, experiment with the music composed of simple and difficult music or including adlib part will be conducted. Moreover, with the results of experiment, model of cooperative performance will be developed.

References

1. Seashore, C.E.: Psychology of music. McGraw-Hill, New York (1938)
2. Todd, N.P.: A model of expressive timing in tonal music. Music Perception 3(1), 33–58 (1985)
3. Shaffer, L.H., Clarke, E., Todd, N.: Metre and rhythm in piano playing. Cognition 20, 61–77 (1985)
4. Okudaira, K., Hirata, K., Katayose, H.: Relationship Between Groove Feeling and the Timing and Loudness of Drum Attacks in Popular Music (in Japanease). IPSJ SIG Technical Report, 2004-MUS-56, vol. 2004(84), pp. 21–26 (2004)
5. Watanabe, T., Chikayama, T.: Analysis of Groove Feeling of Drums Plays (in Japanease). IPSJ SIG Technical Report, 2006-MUS-67, vol. 2006(113), pp. 27–32 (2006)
6. Yamamoto, T., Miyake, Y.: Hierarchical control system of musical performance and design of musical interaction system. In: Proc. of HCII 2005, CD-ROM (2005)
7. Goebl, W., Palmer, C.: Synchronization of Timing and Motion Among Performing Musicians. Music Perception 26(5), 427–438 (2009)
8. Katahira, K., Nakamura, T., Kawase, S., Yasuda, S., Shoda, H., Draguna, M.R.: The Role of Body Movement in Co-Performers' Temporal coordination. In: Proc. of ICoMCS, pp. 72–75 (2007)
9. Horiuch, Y., Sakamoto, K., Ichikawa, A.: Estimating the Model of Human Ensemble (in Japanease). IPSJ Journal 45(3), 690–695 (2004)
10. Kobayashi, Y., Miyake, Y.: New ensemble system based on mutual entrainment. In: Proc. of 12th IEEE Int. Workshop on Robot and Human Interactive Communication, Silicon Valley, U.S.A, vol. 235, pp. 1–6 (2003)

GMM Parameter Estimation by Means of EM and Genetic Algorithms

Sergey Zablotskiy, Teerat Pitakrat, Kseniya Zablotskaya, and Wolfgang Minker

Department of Information Technology
University of Ulm, Germany
{segey.zablotskiy,teerat.pitakrat,
kseniya.zablotskaya,wolfgang.minker}@uni-ulm.de

Abstract. Most of the state-of-the-art speech recognition systems use Hidden Markov Models as an acoustic model, since there is a powerful Expectation-Maximization algorithm for its training. One of the important components of the continuous HMM we focus on is an emission probability which can be approximated by the weighted sum of Gaussians. Although, EM is a very fast iterative algorithm it can only guarantee a convergence to a local result. Therefore, the initialization process determines the final result. We suggested here two modifications of genetic algorithms for the initialization of EM. They are compared to the results of the EM with the same number of local multi-starts.

Keywords: Hidden Markov Model, Gaussian Mixture Model, Expectation-Maximization.

1 Introduction

The Hidden Markov Models (HMM) are commonly exploited in most speech recognition systems [1] because of their efficiency and the existence of a powerful algorithm for their learning - the Baum-Welch algorithm [2]. It is a kind of the Expectation-Maximization (EM) algorithm specially designed for HMM parameters estimation.

Acoustic features extracted from an audio signal (observations in terms of HMM) have naturally a continuous distribution. The use of vector codebooks simplifies the problem of HMM training by transforming the continuous space into vectors from a codebook. However, such conversion decreases the quality of an acoustic model and consequently the performance of the whole speech recognition system. The use of continuous HMMs is therefore strongly preferable in spite of their complexity.

One of the important HMM components we are going to focus on is an emission probability density for each state (normally, one part of a phoneme). According to the central limit theorem [3] whenever the observation is the sum of a large number of random mutually independent events having the same distribution with its own mean and variance it tends to be normally distributed. All the random variables should have approximately equal distribution scale. This implies that there should not be any dominating random variables between them. Unfortunately, this is not mostly the case when the data are coming from different sources. Therefore, for better approximation

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 527–536, 2011.
© Springer-Verlag Berlin Heidelberg 2011

of the multivariate acoustic feature density distribution it is common practice to use a weighted sum of Gaussians (GMM) instead.

EM is a very fast iterative algorithm commonly used for GMM parameters (weights, means and covariance matrices) estimation [4]. However, it can only guarantee a convergence to a local result [5] - local maximum of the log-likelihood. The EM algorithm (as any other local optimization algorithms) strongly depends on the starting procedure. Any unsuccessfully initialized local optimization algorithm gets stuck in a sub-optimal solution, which may be significantly worse than the existing best (global) one.

Quite a standard approach to get a better modeling result is the multi-start from different initial points in the search space. The use of the stochasticity is the only known acceptable way to find the global (or at least better than one of the local solutions) extremum (we do not take into account the examination of the whole search space with a certain small step, because it needs an extremely huge computational power for a multi-dimensional space).

There is a variety of stochastic algorithms for global optimization. However, all of them find the best solution in a local extremum area not faster and not more precisely than the most local search algorithms (gradient or direct search-for-optimum methods).

Since EM is a very powerful and fast local optimization algorithm for HMM parameters estimation it is worthy to use it in a combination with a global stochastic algorithm for its initialization.

In this paper we therefore describe a hybrid approach for the GMM training based on a global optimization genetic algorithm, which generates a new population of starting points for the local EM.

The paper is organized as follows. Sections 2 and 3 outline the use of the EM algorithm for GMM parameters estimation. Section 4 explains the standard genetic algorithm and Section 5 describes its implementation for EM algorithm initialization. Section 6 presents the results. Finally, the conclusions are made in Section 7.

2 The EM Algorithm for GMM Parameters Estimation

The common idea of the EM local optimization method [4] can be described in the following way. At first, the starting point for the algorithm is chosen randomly or according to some a priori information. Then, the iterative process is performed (the recursive formulas are specially designed for different tasks). At each iteration it finds a new function which is a lower bound for an objective function anywhere in the neighborhood of the current point and the value of this function is equal to the value of the objective function in the same current point. For the EM algorithm there is a convergence guarantee to a local solution or at least that it doesn't go downhill. The success of the EM algorithm depends on the objective function. Sometimes it is very hard to find such a lower bound function, but for the density estimation problem there is a good practical solution, which makes the idea of EM work very well [6]. Moreover, one of the important properties of the EM algorithm for density parameter estimation is that it guarantees, in contrast to many other optimization algorithms, an increase of the likelihood function at each iteration, but only until a local maximum is reached.

GMM is a probabilistic model for density estimation using the Gaussian mixture distribution:

$$f(\mathbf{x} \mid \Theta) = \sum_{k=1}^{M} \alpha_k N(\mathbf{x}, \mu_k, \Sigma_k), \tag{1}$$

where

$\Theta = \{\alpha_k, \mu_k, \Sigma_k\}$, $k = 1, \ldots, M$ are parameters of GMM to be estimated,

α_k are Gaussian mixture weights $(\sum_{k=1}^{M} \alpha_k = 1)$,

μ_k, Σ_k are the mean vectors and covariance matrices,

M is the number of mixture components,

$N(\mathbf{x}, \mu_k, \Sigma_k)$ is the probability density function of the *d*-dimensional Gaussian distribution:

$$N(\mathbf{x}, \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{d}{2}} \left|\Sigma_k\right|^{\frac{1}{2}}} \exp\left(\frac{1}{2}(\mathbf{x} - \mu_k)^T \Sigma_k^{-1} (\mathbf{x} - \mu_k)\right). \tag{2}$$

The new parameters Θ^{new} are estimated in terms of old parameters Θ^{g} as follows:

$$\alpha_k^{new} = \frac{1}{N} \sum_{i=1}^{N} p(k \mid \mathbf{x}_i, \Theta^g), \tag{3}$$

$$\mu_k^{new} = \frac{\sum_{i=1}^{N} \mathbf{x}_i p(k \mid \mathbf{x}_i, \Theta^g)}{\sum_{i=1}^{N} p(k \mid \mathbf{x}_i, \Theta^g)}, \tag{4}$$

$$\Sigma_k^{new} = \frac{\sum_{i=1}^{N} p(k \mid \mathbf{x}_i, \Theta^g)(\mathbf{x}_i - \mu_k^{new})(\mathbf{x}_i - \mu_k^{new})^T}{\sum_{i=1}^{N} p(k \mid \mathbf{x}_i, \Theta^g)}, \tag{5}$$

where

N is the training set size,

$p(k \mid \mathbf{x}_i, \Theta^g)$ is a probability of data sample $\mathbf{x}_i$ to be emitted from the k^{th} mixture component with parameters Θ^g :

$$p(k \mid \mathbf{x}_i, \Theta^g) = \frac{\alpha_k^g N(\mathbf{x}_i, \mu_k^g, \Sigma_k^g)}{\sum_{l=1}^{M} \alpha_l^g N(\mathbf{x}_i, \mu_l^g, \Sigma_l^g)}. \tag{6}$$

The above formulas perform expectation and maximization steps of EM at the same time. The results from the previous iteration become guess parameters for the next iteration. The algorithm repeats these steps until a stop condition is satisfied.

3 Initialization of the EM Algorithm

Quite a trivial approach for the initialization of parameters Θ^g is a generation of the uniformly distributed random samples in the range of the available data. However, it often leads to the underflow problem (the logarithm is calculated for a number close to 0) or simply does not find a good result.

Therefore, we used another initialization method. We will further refer to this method as a "local initialization".

The guess values for all the weights are equal to: $\alpha_k^g = \frac{1}{M}$, where $k = 1, \ldots, M$, M is the number of mixture components.

M initial mean vectors are randomly chosen without repeating and modification from the available data set for all the Gaussian elements in the mixture. Suppose that we have an observation sequence

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nd} \end{bmatrix}, \tag{7}$$

where d is the number of dimensions of the mixture model and N is the number of generated data from each dimension, the mean vectors are chosen by generating a set of random values $R = \{r_1, \ldots, r_M\}, r_k \in [1, N]$:

$$\mu^g = \begin{bmatrix} x_{r_1 1} & x_{r_1 2} & \cdots & x_{r_1 d} \\ x_{r_2 1} & x_{r_2 2} & \cdots & x_{r_2 d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{r_M 1} & x_{r_M 2} & \cdots & x_{r_M d} \end{bmatrix}. \tag{8}$$

All covariance matrices are the same diagonal with non-zero elements obtained by estimation of the variances from the data separately for each dimension:

$$\Sigma^g = \begin{bmatrix} \sigma^2(x_{i1}) & 0 & \cdots & 0 \\ 0 & \sigma^2(x_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2(x_{iD}) \end{bmatrix}, \tag{9}$$

where $i = 1,\ldots N$. This covariance matrix is used as an initial parameter for all the Gaussian elements in the mixture model.

4 Genetic Algorithm

A genetic algorithm [7] is aimed to mimic the evolution of the nature. We assume that the parameters of the problem can be encoded into chromosomes. Each chromosome represents one individual which adapts and evolves in a given environment. Stronger and better individuals tend to survive and reproduce the next generation. With the help of crossover and mutation in the reproduction process, chromosomes of the offspring inherit partially the properties of both parents and get new properties. The nature-like process will then select a better offspring for the next generation. This makes the species better and better over time.

In order to utilize the genetic algorithm, the encoding scheme of the chromosomes must be chosen. The classical encoding method is to encode them into a binary string. Each parameter is converted to a binary form and placed into a specific position in the long chromosome. Fig. 1 shows an example of two chromosomes that are encoded into a binary string, where p_i , $i = 1,\ldots n$ are parameters of the problem. All of them are encoded with a 4-bit binary code.

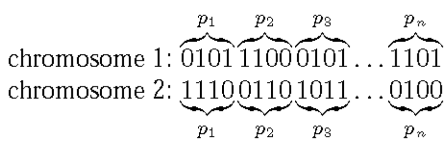


Fig. 1. Chromosome Encoding in binary

Advantages of this method are its simplicity and traceability. Gray code can also be applied to this type of encoding which makes Hamming distance between each value constant. One of the drawbacks of the binary encoding is the encoding time. All parameters must be converted from real values to binary, processed in the evolution and converted back to real values again to be evaluated for the fitness. This conversion takes time and slows down the algorithm when used with very long chromosomes.

The second drawback is the resolution of real values when converted back from binary. As a binary number is discrete, it cannot represent the precision of a floating point in real values. Hence, the number of bits needs to be considered. With an increasing number of bits, the accuracy increases as well. However, the speed mentioned previously and the memory consumption must be taken into account as most compilers allocate the same amount of memory for one bit and one character.

Another method is to use real values directly as parameters of the algorithm (see Fig. 2). This method draws considerable interest since there is no base conversion, therefore the speed is improved. However [8] states that a real-value genetic algorithm may not always give good results depending on the application.

$$\begin{array}{l} \text{chromosome 1: } \overbrace{3.2}^{p_1}, \overbrace{2.5}^{p_2}, \overbrace{7.8}^{p_3}, \ldots, \overbrace{-5.0}^{p_n} \\ \text{chromosome 2: } \underbrace{4.8}_{p_1}, \underbrace{-1.2}_{p_2}, \underbrace{0.4}_{p_3}, \ldots, \underbrace{1.7}_{p_n} \end{array}$$

Fig. 2. Chromosome Encoding in real values

In this experiment, we use a binary encoding scheme for the parameters of each Gaussian element which are weights (α), means (μ) and covariance matrices (∑). These three parameters are converted to binary and concatenated to form a binary string as shown in Fig. 3.

α_k	$\mu_{k1}, \mu_{k2}, \ldots, \mu_{kd}$	$\sigma_{k_{11}}, \sigma_{k_{12}}, \ldots, \sigma_{k_{dd}}$

Fig. 3. Encoded parameters of one Gaussian element

Binary strings from all the Gaussians are again concatenated into a long chromosome of the Gaussian Mixture Model that will be used in the evolution process of the genetic algorithm.

Gaussian 1.	Gaussian 2.	...	Gaussian M.

Fig. 4. Encoded chromosome containing all elements of a Gaussian Mixture Model

During the evolution process chromosomes have to pass through three steps which are selection, crossover and mutation. The selection process chooses a number of chromosomes in the current generation to produce offspring. The selection criterion is primarily based on the fitness of each chromosome. Fitter individuals are more likely to be chosen for reproduction.

Crossover or recombination is a very important process that allows the chromosomes of the new offspring to have some parts of both parents. This is based on the assumption that chromosomes of the parents may contain both good and bad parts. If chromosomes of the new offspring can take some good parts from the first parent and some good parts from the second parent, they tend to become better. Crossover is a probabilistic method that may result in a worse chromosome than those of the parents. The fitness function handles it by biasing good chromosomes so that they are more likely to be selected for the next generation.

Mutation improves the genetic algorithm by altering some parts of the chromosomes that prevents the algorithm from getting stuck in a local extremum. The mutation rate should be properly determined (automatically or through experiments) since too small value may lead to the early local convergence and too large value turns the algorithm into a pure random search. Mutation causes a flip of each bit of the chromosome from zero to one and vice versa with probability equal to the mutation rate.

5 EM Multi-Start by Means of a Genetic Algorithm

The standard genetic algorithm is a global stochastic search which looks for the best solution in the whole search space. However, for high-dimensional tasks where the search space is extremely large it may fail to converge to the global solution in a limited time frame. This is called a genetic drift problem [9] and caused by the accumulation of stochastic errors due to a finite population. The problem occurs when the best solution lies in the vicinity of the found solutions but the genetic algorithm misses or fails to pick up this point.

Since the EM algorithm is very fast and powerful but is only able to find a local extremum in the area of the starting point, we decided to use a genetic algorithm for the initialization as it tends to find the global solution due to its stochastic evolutionary nature. The hybrid EM and GA method is described as follows:

```
Algorithm
  Initialize (Population);
  Evaluate_EM (Population);
  While (Generation no. < Max Number)
        Selection;
        Crossover;
        Mutation;
        Reinsertion;
        Evaluate_EM (New population);
  End while;
End
```

The first generation of GA is initialized by the method described in Section 3. Each individual (see Fig. 3 and 4) is decoded and used as an initial parameter for the Gaussian mixture model. The EM algorithm iterates until convergence to a local maximum and the found likelihood is assigned to the chromosome's fitness value. GA then performs its normal operations by creating a new generation taking into account the fitness value of each individual. These steps are continued until the maximum number of generations is reached.

6 Experiment Results

The data for the experiments (observations) were generated from a given Gaussian mixture model. These parameters were not known to the optimization algorithms and were used only at the generation step. The task was to estimate the parameters of GMM, given the data set.

The algorithms investigated in this work were multi-start EM with local initialization (see Section 3) and two hybrid Genetic Algorithms in combination with EM: randomly initialized GA and locally initialized one.

As the following step, the multi-start EM algorithm was performed. The starting point was determined by the local initialization. The algorithm iterated until a convergence to a local maximum of the likelihood was reached. Then, the best result was saved and a new initial model was again initialized locally, independently from the previous initialization and results. Finally, the best model with the highest likelihood among these multi-starts was accepted as the solution. The randomly initialized GA created the first generation randomly in the search space, regardless of the available data. Each individual (initial parameters of GMM) was passed to the EM algorithm. It then performed a standard procedure to find a local maximum of this model. The resulted likelihood was assigned to the individual as the fitness value. The next generation was a product of selection of the parent individuals, their crossover and mutation. The best individual in the last generation was then chosen as the best found model (taking into account elitism property).

Locally initialized GA performed the evaluation process with the help of EM in a similar way as random initialized GA did. The only difference was the initialization of the first GA generation. Individuals were chosen from the data by the local initialization method (the same with the EM Multi-Start initialization).

All three algorithms were set to run 1000 times of EM restarts. That is, for the standard EM multi-start, 1000 restarts were performed. For genetic modifications, where each individual was equal to one EM restart, the total number of individuals (number of individuals per generation × number of generations) had to be 1000. Furthermore, since the initialization process in EM and GA operations were all stochastic, the final results were averaged by 100 same experiments.

Fig. 5 shows the simulation results for a case, when the real number of components in the mixture was known or correctly determined (by clustering or as a priori information) and equal to 6. The dimension of the task was chosen to be 39 similarly to the dimension of SR task. The audio features extracted from a speech signal are usually the energy and 12 Mel-frequency cepstrum coefficients [10], their first and second order derivatives, i.e. 39-dimensional feature vectors.

As can be seen, the locally initialized GA performed better than randomly initialized one (for both cases: 20 and 50 individuals per generation). Locally initialized GA gave a better result (smaller value of the negative log-likelihood) than the standard EM multi-start after 200-250 restarts (the total number of individuals).

Fig. 6 shows the simulation results for a case, when the real number of components in the mixture was not known. The data was generated from GMM, consisting of 6 components and modeled by a mixture with 4 components. This is quite a standard situation for speech recognition tools (and not only) that the number of Gaussians in the model differs from the real data distribution (because of a lack of knowledge). It can be seen that both GA modifications performed significantly better than EM multi-start from the very beginning.

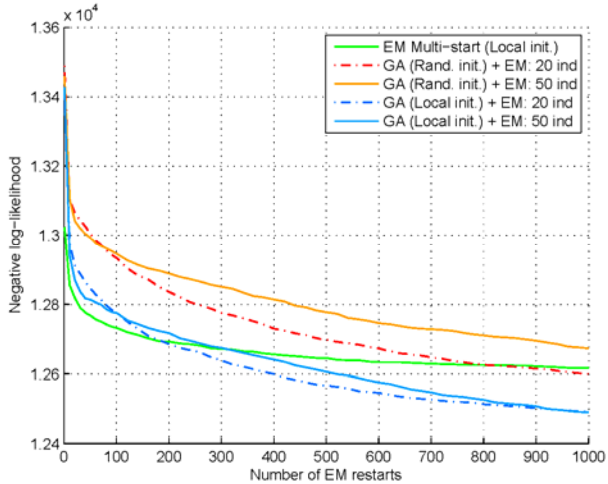


Fig. 5. EM Multi-start vs. Hybrid GA: 6 Mixture Components modeled by 6 Components

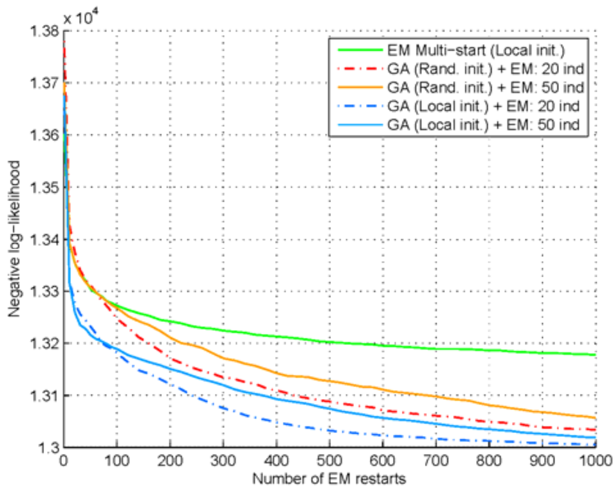


Fig. 6. EM Multi-start vs. Hybrid GA: 6 Mixture Components modeled by 4 Components

In these two simulations, a small mutation rate 0.005 (for each bit) was used which gave better results than higher mutation rates, 0.01 and 0.05.

Other types of selection and crossover were also investigated and the best methods were stochastic universal sampling and double-point crossover. However, they did not have significant influence on the likelihood value compared to the mutation rate.

The number of individuals also influenced the result. In most cases 20 individuals per generation (50 generations) led to a better result than 50 individuals (20 generations). A larger number of individuals in one population increases the chances of the algorithm to get closer to the global minimum at the initialization step. However, at the same time it reduces the number of GA iterations for processing these starting points. Hence, the achieved result means that the GA operators outperform the pure EM algorithm restart from randomly chosen points.

7 Conclusion

The employment of the genetic algorithms for the initialization of the Expectation-Maximization algorithm mostly (and significantly) increases the likelihood of the data being generated by the found Gaussian Mixture Model. It is shown that the local initialization allows us to avoid the underflow problem and leads usually to a better result.

However, the use of genetic algorithms is worth only if there is an opportunity (computational power and time) to restart EM algorithm for a certain number of times. It does not bring us any significant effect when only few generations of GA can be performed.

References

1. Peinado, A., Segura, J.C.: Speech Recognition over Digital Channels: Robustness and Standards. John Wiley and Sons Ltd, Chichester (2006)
2. Rabiner, L., Juang, B.-H.: Fundamentals of Speech Recognition. Prentice-Hall, Englewood Cliffs (1993)
3. Rice, J.: Mathematical Statistics and Data Analysis, 2nd edn. Duxbury Press, Boston (1995)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39(1), 1–38 (1977)
5. McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions. Wiley, New York (1997)
6. Redner, R., Walker, H.: Mixture densities, maximum likelihood and the EM algorithm. Society for Industrial and Applied Mathematics (SIAM) 26(2), 195–239 (1984)
7. Davis, L.: Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
8. Goldberg, D.E.: Real-coded genetic algorithms, virtual alphabets, and block. University of Illinois, Urbana (1990)
9. Nolle, L., Battersby, A., El-Mihoub, T.A., Hopgood, A.A.: Hybrid Genetic Algorithms: A Review (2006)
10. Davis, S.B., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing 28(4) (1980)

Part VI
Image Processing and Retrieval in HCI

Shape –Based Human Actions Recognition in Videos

Nitish Amraji[1], Lin Mu[2], and Mariofanna Milanova[1]

[1] Computer Science Department, University of Arkansas at Little Rock, USA
[2] Computational Science Department, University of Arkansas at Little Rock, USA
{nxamraji,lxmu,mgmilanova}@ualr.edu

Abstract. The paper presents a system for human action recognition using contour based shape representation. With the rapid progress of computing and communication technology smart user computer interfaces are becoming most widespread. A major goal is to go further than traditional human computer interaction (like mouse or keyboard) and to find more natural means of interaction with computers, including the application of computer games and surveillance. The objective of this work is to achieve representation eigenspace for modeling and classifying actions performed by individuals. Eigenspace is the subspace for each type of action. A representation eigenspace approach based on the Principal Component Analysis (PCA) algorithm is used to train the classifier. Behaviors are classified with respect to a predefined set of learning actions. The key points of this approach include the mode silhouettes are extracted from video, the kind of shape descriptor used, the development of the new eigenspace and the kind of classification used. Performance of the system is expressed in terms of percentage of right or wrong classifications.

Keywords: Human Activity Recognition, Fourier Descriptor, Principal Component Analysis.

1 Introduction

Human activity recognition has recently gained growing interest from the computer vision research community. Previous work on human action recognition can be classified in two main categories: Motion –based analysis and shape- based analysis. Motion – based approaches extract human movement character information to represent human activity such as movement direction, body parts position, velocity, optical flow, etc. Shape-based approaches use geometric information such as silhouette, contour and deformed shape to represent human activity. In this paper, we propose to use the shape – based approach. The key points of these applications are: how the silhouettes are extracted and what kind of shape descriptor is used. Silhouettes and edges are used the most because they can be easily extracted and are to some extent, lighting invariant. Active contours are often used in computer vision and image analysis to detect and locate objects, and to describe their shapes. The general disadvantages of these models are that the active contour may leak out of the ideal contour when the edges are weak and in many cases its extraction is very difficult.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 539–546, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Since the active contour model is edge-based, the dynamic active contour captures the sharp edge of the shadow. Other algorithms for contours extraction are based on various techniques [1]. For the contours presentation different shape descriptors have been developed. Many authors proposed to use Fourier descriptor (FD) for shape analysis, and shape coding. The main advantages of their use are that FD not only overcomes the weak discrimination ability of the moment descriptors and the global descriptors, but that FD also overcomes the noise sensitivity [2]. In this work, the FD was applied on the human body silhouettes represented by their main contours.

The advantages of FDs are that they achieve both good representation and normalization, they are compact, can be calculated efficiently, they are invariant in terms of position, rotation and scale and are to some extent insensitive to local noise.

1.1 Overview of the Proposed System

In this study we developed an automatic human action recognition system based upon spatiotemporal silhouette analysis measured during different activities. Intuitively, recognition of human action from video depends on how the silhouette shape of on individual changes over time in an image sequence. We may consider human action recognition to be composed of a sequence of static body poses and expect that some distinguishable signatures related to these static body poses can be extracted and used for action recognition. Many researchers proposed to use similar ideas for face recognition or gait recognition using silhouette representation and eigenspace transformation using PCA [3]. The overview of the proposed system is shown in Fig. 1.

The proposed approach includes first, the extraction of curve-based shape descriptors, such as Fourier descriptors. Second, for each monitor activity, a PCA is implemented for dimensionality reduction and incremental classifier training and third the matching and classification. The system learned to classify five different actions performed by a single subject and recorded as an image sequence.

1.2 Purpose and Contribution of the Paper

The main purpose and contributions of the current paper are summarized as follows: We attempt to develop an effective system for human actions recognition using silhouette signature representation. We apply the PCA method like [3] to reduce the dimensionality of the input feature space

The body contours extraction is performed using two consecutive filters: the first - for the adaptive texture suppression, and the second – for the contours extraction. Compared with most contour extracting algorithms our approach gives a complete silhouette representation. We have reduced search space first time implementing body contour signature representation and second time implementing PCA.

The paper is arranged as follows: in Section 2 is explained the method used for the contours extraction, in Section 3 is presented the action recognition module , in Section 4 are given some experimental results, and Section 5 is the Conclusion.

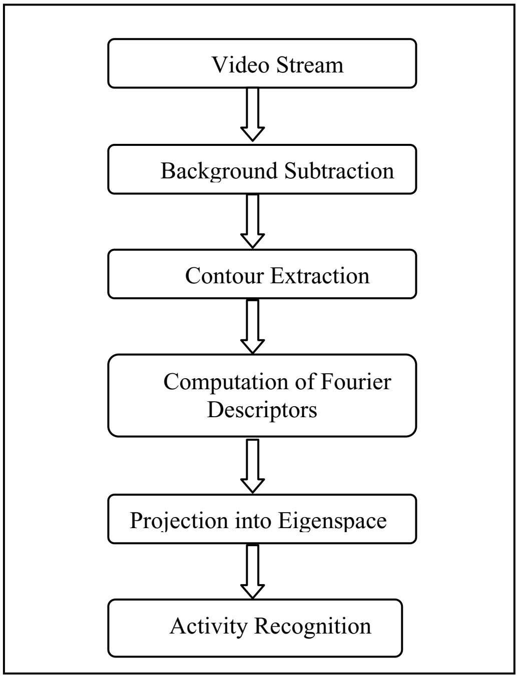


Fig. 1. Illustration of Processing stages of our approach

2 Body Contour Signatures Representation

2.1 Body Contour Extraction

Human tracking and body contour segmentation are basic and important steps in our system. The action recognition final result depends greatly on the quality of contour segmentation.

We have proposed human body movements tracking system using video imagery. The system contains the following sub modules: person detection, subject tracking and silhouette extraction [4].

In the first step, each frame of the video sequence is processed to extract the Region of Interest (ROI) from the background. For each frame of the flow we compute a set of scalars that characterizes the shape of the flow in that frame. We use all the points in the flow and analyze their spatial distribution. The shape of motion is the distribution of flow, characterized by several sets of measures of the flow. For example, we compute the following scalars: x and y coordinates of the centroid of moving region, aspect ratio of moving region. The extracted ROI is then filtered to evaluate a vector of feature maps which is used to represent the spatial distribution of features in the frame. The detected blobs are refined to produce a human silhouette.

The extraction of the main contours in the image is based on image segmentation. The aim is to define the parts, whose brightness is lower and higher than that of the already equalized image brightness. For result of the processing described above, the histogram has only one maximum, which corresponds to the equalized object brightness and the detection of other maximums, which to indicate other significant parts or objects, is not possible. For this reason, the image segmentation is based on the so-called “triangle” algorithm [5], modified for this application. The modification concerns the faster detection of the segmentation thresholds values .The proposed contour segmentation algorithm is presented in [1]. A comparison with other algorithms for contour segmentation demonstrates that the continuous structure of the contours extracted with the new method, allows for the representation of a much higher numbers of details.

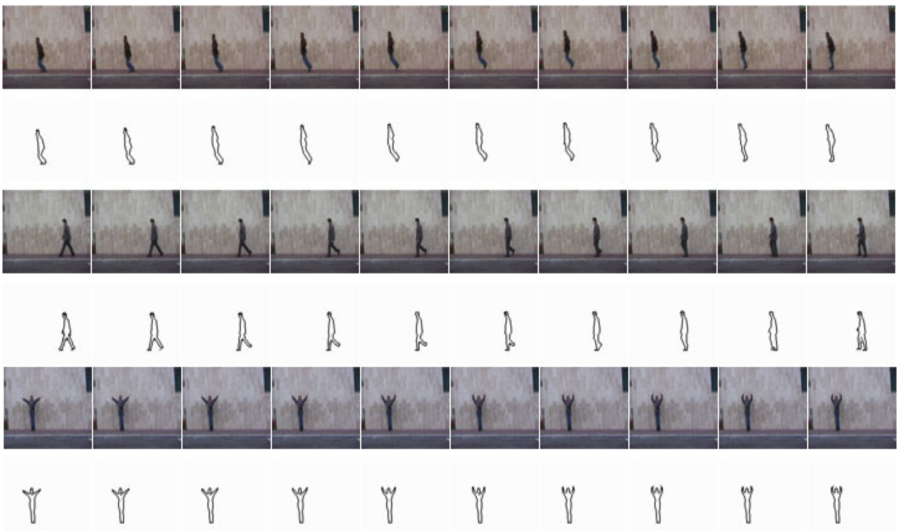


Fig. 2. Examples of video sequences and extracted contours from database

2.2 Fourier Descriptors and Contour Signatures

Many authors propose to use curve-based shape descriptors such as the Fourier descriptor, (FD) for the contours representation. The FD advantages are that they achieve both good representation and normalization; they are compact, can be calculated efficiently; they are position, rotation and scale invariant and are to some extent insensitive to local noise. FDs inherently capture perceptual shape characteristics and their lower frequencies correspond to the average shape, while higher frequencies describe shape details. FDs can be compared with a Euclidian distance. The main idea of Fourier descriptors is to describe a silhouette by fixed number of sample boundary points (x0, y0, …, xk, yk). The points are sampled using equidistance sampling, transformed into complex coordinates and are further transformed to the frequency domain using a Discrete Fourier Transform (DFT). The results of this transformation are called the Fourier coefficients, denoted by (f0, f1,…,fk-1). Since f0 is always zero and f1 is always one, we have k-2 unique coefficients. The DC component f1 only indicates the shape position. The FDs are given by FD = (|f2|/|f1|, |f3|/|f1|, ..., |fk-1|/|f1|). This descriptor is a shape signature and can be used as a basis for similarity evaluation and shape body retrieval.

Fourier Descriptors of boundaries are the complex coefficients of the discrete Fourier transform (DFT) with the coordinate pair on the boundary point, which was chosen to show the images silhouettes. An effective way to capture the gross essence of the boundary is by choosing different number of Fourier Descriptor coefficients. And thus we can use that as the basis for differentiating between distinct shapes.

We have plotted the magnitude of the Fourier Descriptor for several human silhouettes, we found out that of all the FD's of silhouettes only few terms have higher magnitude and give the global form of the silhouette. So we have fixed the number of FD's for all the images in our dataset. We only used 60 descriptors to characterize the human silhouette. Fig. 3 shows an example of human silhouette representation using all the Fourier Descriptor terms and using 60 FD terms.

Fig. 3. Human Silhouette Representation. (a) Is the silhouette representation using all the Fourier Descriptor terms (i.e. 200 terms for the given example) and (b) is the silhouette reconstruction using 60 terms.

The correspondent contour signature in each frame is grouped with the previous computed signature to generate an image that is related to the action performed. The dimension of the image depends on the temporal duration of the action. Assuming the temporal duration of the action is in L. Then the resulting contour signature images will be of size L x M. The contour image is given by

$$C_k = \begin{bmatrix} F_{k,1} & F_{k.2} & . \; . \; . & F_{k,M} \\ F_{k+1,1} & F_{k+1,2} & . \; . \; . & F_{k+1,M} \\ . & . & . \; . \; . & . \quad .. \\ . & . & . \; . \; . & . \\ F_{k+L-1,1} & F_{k+L-1,2} & . \; . \; . & F_{k+L-1,M} \end{bmatrix} \tag{1}$$

The row $\{F_{k,1}, F_{k,2}, \ldots F_{k,M}\}$ is the Fourier descriptor of the contour signature in kth frame. Matrix Ck is called feature image related to one activity

3 Action Recognition

The main goal of action recognition module is to classify actions into one of several categories. As the input for this module we use the feature image representation calculated throughout the action duration. We compare the query feature image with reference feature images of different learned actions and look for the best match. There are several issues to consider using this approach. First action duration is not fixed for different actions. Second we need to transform high-dimensional feature space to low-dimensional eigenspace. We propose to use Principle Component and represent an action by s single point in eigenspace. The advantages of doing PCA is: reducing search complexity and increasing robustness

The projection into eigenspace is computed as

$$Y_{i,j} = \phi^T C_k = \begin{bmatrix} y_1^{ij} & y_2^{ij} \ldots\ldots y_{T_{ij}}^{ij}. \end{bmatrix}. \tag{2}$$

Each Y_k^{ij} is an m-dimensional column vector which represents a point in eigenspace. Y_k^{ij} is a manifold representing a simple action?

The recognition step is performed by comparing the manifold of new test action to the train reference manifolds. Usually motion sequences have different lengths (number of frames) and ideal distance metric should be able to handle such changes. To compare the distance between two manifolds we are using the distance measure variant of the Hausdorff metric. Our distance measure can handle changes in duration and is invariant to temporal shifts.

Given two manifolds from two activity $A = [a_1.a_2 \ldots\ldots a_n]$ and $B = [b_1.b_2 \ldots b_m]$, we define

$$d(A,B) = \frac{1}{n}.\sum_{j=1}^{n} \min \left\| \frac{a_j}{\|a\|} - \frac{b_i}{\|b_i\|} \right\| \tag{3}$$

To ensure similarity, we use the following distance measure presented in [7].

$$D(A,B) = d(A,B) + d(B,A) \tag{4}$$

We adapt the Nearest Neighbor classifier (NN) for final activity classification. Let T represent a test sequence and Ri represent the reference sequence of class I, we may classify this test sequence into class I that can minimize the similarity distance between the test sequence and the reference sequence

$$c = \arg\min D_i(T, R_i) \tag{5}$$

Where D is similarity measure describe in (4)

4 Experiments

We use the action database [6] to classify five types of activity: walking, running, jump, one-hand wave, two-hand wave. At first, we extract sequences from the original videos performed by 9 different people. And then we use the FD described in Section 3 to describe the contour image of the human body in each frame. The number of Fourier Descriptor terms used for all the frames was fixed which give the global form of silhouette (i.e. 60 terms in our case). We use the methods described in Section 5 to perform activity recognition. The sequences were divided into training set (2-4 persons) and a test set (7-5) for different experiments. The results of the classification are shown as a confusion matrix in Table 1 when 2 persons were used for training set and remaining 7 persons were used for test set. The table displays the number of sequences classified correctly in the diagonal. The average recognition rate obtained for all the three experiments was around 86%.

Table 1. Confusion Matrix for the experiment

Action	**Walk**	**Run**	**Jump**	**One-Hand Wave**	**Two-Hands Wave**
Walk	6	1			
Run	3	4			
Jump			7		
One-Hand Wave				5	2
Two-Hands Wave				1	6

5 Conclusions

Automatically recovering human actions from images is useful but challenging problem due to the multiple parameter variations (changing illumination, shadows, etc.). In this paper, we present a framework for action recognition using Fourier descriptors and eigenspace representation. The contour extraction of moving human subjects is performed using adaptive contours extraction. Then, Fourier descriptors with a fixed number of sample points are used to describe the extracted contour silhouettes and finally similarity distances are measured. Experiments with varying number of sampled points are performed on test data, containing random positions and body movements. Performance of the system is expressed in terms of percentage of right or wrong classification.

Acknowledgments. This paper was supported by the System Research and Application (SRA) Contract # 0619069.

References

1. Mendi, E., Milanova, M., Kountchev, R., Kountcheva, R., Todorov, V.: Extraction of Human Body Contours and Position Analysis. In: CISP 2010, pp. 2765–2769 (2010)
2. Ling, Z., Zhao, C., Pan, Q., Wang, Y., Cheng, Y.: Analyzing Human Movements from Silhouettes via Fourier Descriptor. In: IEEE Conf. on Automation and Logistics, pp. 211–236 (2007)
3. Wang, L., Tan, T.: Silhouette Analysis – Based Gait recognition for Human Identification. IEEE Transaction on Pattern Analysis and Machine Intelligence 25(12), 1505–1518 (2003)
4. Milanova, M., Bocchi, L.: Video-Based Human Motion Estimation System. In: Duffy, V.G. (ed.) ICDHM 2009. LNCS, vol. 5620, pp. 132–139. Springer, Heidelberg (2009)
5. Young, I., Gerbrands, J., van Vliet, L.: Image Processing Fundamentals. Delft, Netherland (1996)
6. http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html#ClassificationDatabase
7. Papanikolopoulos, M.N.: A method for human action recognition. Image and Vision Computing 21(8), 729–743 (2001)

Dynamic Queries with Relevance Feedback for Content Based Image Retrieval

Murat Birinci*, Esin Guldogan*, and Moncef Gabbouj

Department of Signal Processing, Tampere University of Technology, Tampere, Finland
{murat.birinci,esin.guldogan,moncef.gabbouj}@tut.fi

Abstract. A novel relevance feedback scheme utilizing dynamic queries for content based image retrieval systems is proposed, where the retrieval results are updated instantly based on the user's feedback. The user is expected to label at least one image as positive or negative, revealing the gist of the expected retrieval results. Then the retrieval results are updated dynamically, without any further user interaction, based on the similarity of the query and the selected image in different feature spaces increasing the semantic accuracy of the retrieval. The proposed method not only invalidates the drawbacks of current relevance feedback systems in terms of user experience, but also provides an innovative stand point for the relevance feedback scheme as well.

Keywords: dynamic queries, relevance feedback, content based image retrieval.

1 Introduction

Developments in multimedia technology in the past century altered the media world from analog to digital, resulting in not only putting every individual into a content creator position, but also transferring all the media already at hand into digital format, which significantly increases the amount of digital media available. Visual, aural and textual information can now be created, stored and modified digitally via everyday devices or software such as cameras, audio players, text editors, implying every single user may have a vast amount of such media. Moreover, improvements in network technology, especially in Internet, give rise to distribution of this content over an indefinitely large population, yet bringing forth the problem of management of this information for efficient accessibility. The very first approaches on multimedia indexing utilize text-based annotations, analogous to library catalogues, where a text based description is linked to every database element. However the reliability of such methods is entirely annotator dependent and furthermore, they require vast amount of work especially with increasing database sizes. Such drawbacks are also encountered for any kind of manual indexing method which is why automatic indexing is essential. Automatic schemes intend to overcome any subjectivity and represent the content effectively, defining the scope of *Content Based Multimedia Retrieval* (CBMR).

* These authors contributed equally to this work.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 547–554, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Without any loss of generalization, in this paper we reduce our focus to *Content Based Image Retrieval* (CBIR) which deals with visual features only. Various visual features, such as color, texture and shape, are utilized in CBIR systems in order to describe the content of an image (or any visual media); however their retrieval performance is usually limited especially on large databases due to the lack of discrimination power of such features. Moreover, a more evident reason is while such features strive to extract an objective description of the content, the content is undoubtedly subjective and there is the so called *semantic gap* between the description of the content and its semantic interpretation at the user side. Therefore, incorporation of user subjectivity and experience is of decisive importance in order to achieve competent retrieval results.

Relevance feedback is found to be one of the most powerful methods for improving image retrieval performance and bridging the semantic gap [1]. The word *relevance* in this context is described as "*the ability of a retrieval system to retrieve material that satisfies the needs of the user*". Relevance feedback methods predict and learn user's preference to improve retrieval results. They interact with user during the query in order to get the user's subjective perception for improving the query results via providing the opportunity for evaluating them. It iteratively improves the accuracy of the retrieval by modifying the query, based on the user's feedback. The iterated relevance feedback may guarantee improved retrieval results. However, while being such a supportive and practical tool, most of the relevance feedback methods suffer from user experience side due to its iterative nature. Users are expected to update their queries after each feedback in order the system to learn their behavior and choices. Thus, it is obvious that current CBIR systems are lacking the adjustment to average users' needs, i.e. they cannot synchronize to the way an average user searches, even though the technology is available. Zhang accurately stated the same issue in [2] as "*the user interface design should adapt to users' behavior, not shape users' behavior*". Yet, CBIR society mainly focused on the background technology instead of how to utilize that technology using an appropriate design in order to reach the average user.

The rest of the paper is organized as follows. Section 2 provides an overview of the related work in this area. In section 3, the proposed method is explained in detail and experimental results are given in section 4. Section 5 concludes the paper and suggests possible future work.

2 Related Work

There are various proposed relevance feedback methods in the literature. Most of them treat each iteration as a separate query, i.e. by combining the initial query and the feedback from the user they form a new query and present the reconstructed results [1]. Various feature re-weighting and query re-formulating methods in order to form the new query are introduced in the literature [3], [4]. For example Djordjevic and Izquierdo in [5], used variance in order to describe discrimination power of the feature. Main drawback of such approaches is the necessity of large amount labeled data and several iterations for improving the semantic accuracy. More recent approaches involve neural networks and machine learning approaches [6], [7], in order to learn and model user behavior. However, the computational complexity significantly increases when the learning process is involved. Moreover, it should be noted that the user's preferences might vary in every query. For instance, the user might use the picture in Fig. 1 in order to search for *horses*. However the same picture can be used for searching *beach* or *ocean* another time. Therefore, if the system *learns* that this picture is associated with horses, future searches with this picture will yield results with more horses in them. This is a clear example of what is meant in Section 1 as "shaping users' behavior". However, users' interests and objectives are not the same at all times.

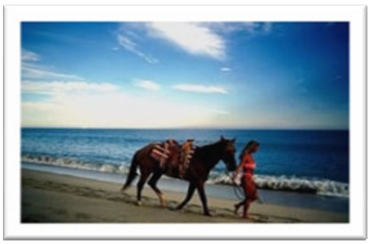


Fig. 1. A picture with various content

An instance-based relevance feedback method is proposed by Giacinto *et al.* [8], where they represent the images in a dissimilarity-space. Their notion of relevance is the degree of similarity of an image to the nearest relevant (based on user feedback) image. Similarly, irrelevance is the dissimilarity from the nearest irrelevant image. Than the images are ranked again according to their relevance score and top K images are displayed. However, while the method points out the significance of query-based feedback instead of an overall learning approach, re-calculation of distances and updating the whole query is no different that re-making the query for the user. Moreover, the number of images expected from the user to mark as relevant or irrelevant is in the order of dozens. Those are again facts that are "shaping users' behavior" instead of "adapting it".

Another instance-based method is proposed in [9], where user feedback is used in re-weighting individual feature distances based on intra- and inter-cluster relations. While this method avoids re-calculation of feature distances (which is computationally expensive and time consuming) and simply re-weights the already calculated distances, it requires an additional query after the feedback using the re-weighted distances. This means re-ranking of *all* images in the database.

Shneiderman presented "direct manipulation" techniques in 1983 [10] and "dynamic queries" in 1994 [11] in order to introduce quick and powerful query methods for database and information retrieval via graphical controls such as sliders, buttons etc. Dynamic queries describe the interactive user control of visual query parameters that generate a rapid animated visual display of database search results. They enable users to conveniently overview, explore and filter out the results and observe the effects immediately. Therefore, the user can obtain the desired results in real-time without any reformulation of the query. In this context, utilization of dynamic queries in relevance feedback methods should be much more than a mere process of gathering relevance information. The feedback should provide a user preference and influence the results instantly.

3 Proposed Method

We propose an instance-based relevance feedback method that utilizes small number of labeled data at that particular query session, i.e. the relevance feedback received from the user affects only the current query and not the future retrievals. The proposed scheme utilizes dynamic querying meaning that the user is not required to repeat or reformulate the query. Instead, the results are updated dynamically based on the feedback.

Users are expected to label the image(s) either positively or negatively according to their subjective preferences. For positively labeled image(s), it is assumed that user wants more images similar to that particular image(s), thus that image(s) will be used for presenting new images to the user dynamically. Similarly, users may negatively label the image(s) for their subjective preferences, meaning that such images are out of user's interest. Based on the negatively labeled images, particular images will be eliminated and new images will be introduced to the user by modeling the user's preferences. Fig. 2 demonstrates how a negatively labeled image updates the current query. Initial retrieval results based on text based search are shown in Fig. 2.a (Yet, we used only visual features in our experiments). User is provided to choose any image as preferred or non-preferred when hovered. In Fig. 2.b, the user does not prefer to have the image in his results and the system removes that image *together with the ones having similar content* (Fig. 2.c) and brings new images from lower ranks not including similar ones to the user's choice (Fig. 2.d).

Fig. 2. An illustration of how negative feedback affects the results

The overall similarity distance between two images is calculated by averaging similarity distances from several different features. If N distinct features are extracted from images I_1 and I_2, than the overall similarity distance D between those two images is calculated as:

$$D(I_1, I_2) = \frac{1}{N}\sum_{i=1}^{N} d_i(I_1, I_2) \tag{1}$$

where $d_i(I_1,I_2)$ is the similarity distance between I_1 and I_2 for the i^{th} feature. However, the fact that relevance feedback is necessary means that the user is not satisfied with the discrimination provided by D. Hence, the user provides his/her preference on which images should be more similar (positively labeled images) and which should be regarded as different (negatively labeled images).

It should be noted that all the labeled images are found to be similar based on the overall similarity distance. Thus, we analyze the feature spaces independently and seek a feature space that provides the requested (dis-)similarity.

Let us consider the positive and negative feedback cases separately. If the user labels an image (namely I_2) as "irrelevant", a feature providing a high similarity distance from the query image (namely I_1) is sought. Therefore we select the i^{th} feature space that provides the maximum similarity distance, i.e. $max(d_i(I_1,I_2))$. In order to remove the images that are similar to I_2 and dis-similar to I_1, we remove the images that have small similarity distance to I_2 in the i^{th} feature space (Fig. 3). The decision on the number of images to be removed can be made in many ways. A distance threshold can be set and images closer than this threshold to I_2 can be removed. This is also illustrated in Fig. 3 where the distance threshold is shown as the green circle around *p2*. In our experiments we used *KNN* (K-Nearest-Neighbors) algorithm in order to remove *K* images nearest to I_2 in the i^{th} feature space.

On the other hand, if an image is labeled as "relevant", we find the i^{th} feature space yielding the minimum similarity distance with the query image and i.e. $min(d_i(I_1,I_2))$. Similarly, the *K* nearest images to I_2 are found and raised to higher ranks. However, while it was easy to simply discard *K* irrelevant images, more analysis is needed in

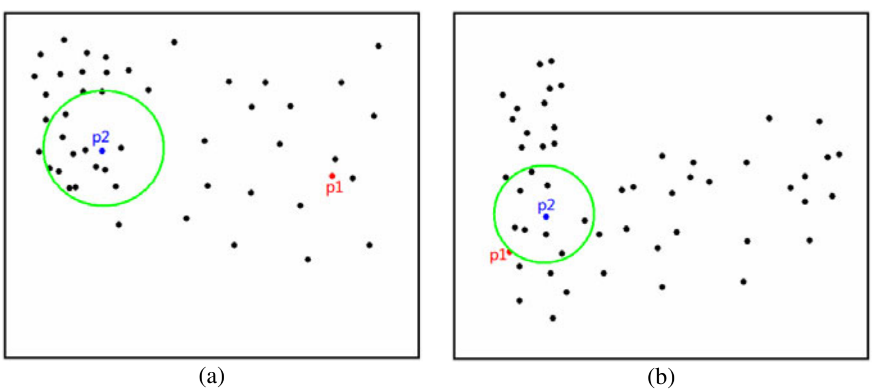


Fig. 3. Two feature spaces that provide different similarity distances between the same elements p1 and p2

order to determine the updated ranks of the *K* relevant images. Incautiously moving the *K* relevant images to the top *K* rank would be interfering with the users' preferences. Thus, in order to include those *K* images among the presented images to the user and determine their appropriate ranks, one should consider both their distances in the i^{th} feature space and the distances of the presented images. Therefore, the distances of *K* relevant images are than re-weighted according to the following formulas:

$$\mu_i = \frac{1}{K}\sum_{m=1}^{K} d_i(I_1, I_m), \tag{2}$$

$$\mu_\forall = \frac{1}{L}\sum_{m=1}^{L} D(I_1, I_m), \tag{3}$$

$$D'(I_1, I_m) = \frac{\mu_i}{\mu_\forall} d_i(I_1, I_m) \tag{4}$$

where *L* is the number of images presented to the user. In other words, in order to map the distances of *K* relevant images to the *L* presented images, their distances are weighted with the ratio of average distances (μi and $\mu_\forall$). The value of *L*, i.e. the number of images presented to the user, depends on the retrieval system and the UI. Hence, in order to promote the *K* relevant images among the presented images, the average similarity distance of the *L* presented images is used to weight the distances of *K* relevant images.

4 Experimental Results

In our experiments, we used Corel real-world image databases for evaluating the retrieval results of the proposed method. Corel image data sets are well-categorized and widely used in the CBIR literature. For evaluating the results, a Corel database with 1000 images are used. These images are pre-assigned by a group of human observers to 10 semantic classes each containing 100 images. The classes are: Africa, Beach, Buildings, Buses, Dinosaurs, Flowers, Elephants, Horses, Food, and Mountains. In our experiments, the following low-level color, shape, and texture features are used: YUV and HSV color histograms with 128-bins, Gray Level Co-Occurrence Matrix texture feature with parameter value 12, Canny Edge Histogram, and Dominant Color with 3 colors. We compared our results with the methods in [8] and [5] which are briefly explained in Section 2. Average precision values are calculated based on the retrieval results from these queries. We used *K*=10 for the calculation of *KNN* as described in Section 3. For our system the number of images presented to the user is 24, i.e. *L*=24.

We performed 17 queries from different classes. Fig. 4 shows the average precision values calculated for the initial query (no user feedback), the proposed method and the competing methods in [8] and [5]. It is clear that the improvement in precision using the proposed method is slightly better than the competing methods. While the performance measures are close to each other, another point that should be considered is the amount of user feedback. Users are not usually in favor of laborious work such as providing feedback. As we discussed in Section 2, the system should not require immoderate user effort, i.e. it should not shape users' actions. While the methods in [8] and [5] require significant amount of feedback (labeled images), the proposed method is designed to work with minimum amount of labeled data (images). Therefore we used 1 positively labeled (relevant) and 1 negatively labeled (irrelevant) image as the user feedback for the proposed method. However, since the competing methods are not designed to work with such limited input, we used 3 positive and 3 negative feedbacks in order to obtain the results in Fig. 4. Even with such a difference in the user input, the proposed method proves to perform better than the competing techniques. Moreover, the computational efficiency of the proposed method outperforms the competing methods. Since other methods require complex computations including calculation of variances etc. our method works 7 times faster than [8] and 50 times faster than [5]. This fact also enables us to utilize dynamic querying by working under the limitations stated in [11].

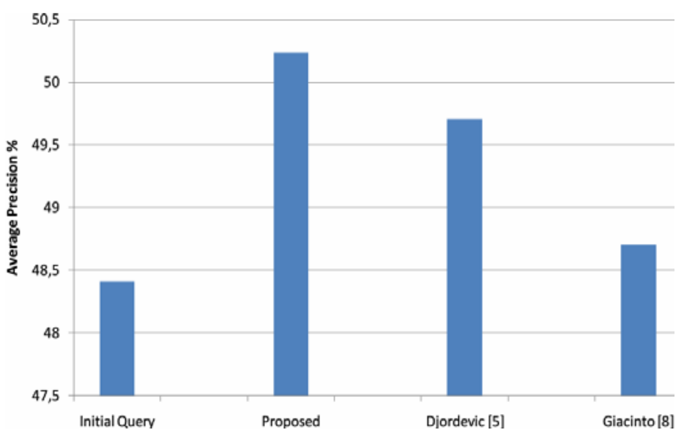


Fig. 4. Performance results for the initial query, the proposed and competing methods

5 Conclusion

We proposed a novel relevance feedback method for content based image retrieval. Dynamic querying techniques are used in order to instantly update the retrieval results. The proposed method treats each query independently, enabling users to have different preferences every time they query. Minimum amount of feedback is required in order to grasp user preference. User feedback is utilized in order to find the feature space that reflects user's view of similarity between the query image and the selected image. Experimental results prove that the proposed method outperforms competing methods both in performance and computational efficiency. The computational efficiency together with dynamic querying and the limited amount of required user labor provides a better user experience than the compared to the state-of-art methods.

References

1. Zhou, X., Huang, T.S.: Relevance Feedback for Image Retrieval: A Comprehensive Review. ACM Multimedia Systems 8, 536–540 (2003)
2. Zhang, D.: A Survey of User Interfaces in Content-based Image Search Engines on the Web. In: Proceedings of the Fourth Workshop on Human-Computer Interaction and Information Retrieval (HCIR), pp. 129–132 (2010)
3. Heesch, D., Rueger, S.: Two step relevance feedback for semantic disambiguation in image retrieval. In: Sebillo, M., Vitiello, G., Schaefer, G. (eds.) VISUAL 2008. LNCS, vol. 5188, pp. 204–215. Springer, Heidelberg (2008)
4. Piras, L., Giacinto, G.: Neighborhood-Based Feature Weighting for Relevance Feedback in Content Based Retrieval. In: Workshop on Image Analysis for Multimedia Interactive Services, London, UK (2009)
5. Djordjevic, D., Izquierdo, E.: An Object- and User-Driven System for Semantic-Based Image Annotation and Retrieval. IEEE Trans. Circuits Syst. Video Techn. 17(3), 313–323 (2007)
6. Rao, Y., Mundur, P., Yesha, Y.: Fuzzy SVM Ensembles for Relevance Feedback in Image Retrieval. In: Proc. of CIVR, pp. 350–359 (2006)
7. Zhang, L., Lin, F., Zhang, B.: Support Vector Machine Learning for Image Retrieval. In: Proc. of ICIP, pp. 721–724 (2001)
8. Giacinto, G., Roli, F.: Instance-based relevance feedback for image retrieval. In: Perner, P. (ed.) Case-Based Reasoning on Images and Signals, pp. 489–496. Springer, Heidelberg (2008)
9. Guldogan, E., Gabbouj, M.: Dynamic Feature Weights with Relevance Feedback in Content-Based Image Retrieval. In: Proceedings of the IEEE International Symposium on Computer and Information Sciences, ISCIS, Cyprus (2009)
10. Shneiderman, B.: Direct Manipulation: A Step Beyond Programming Languages. Computer 16, 57–69 (1983)
11. Shneiderman, B.: Dynamic queries for visual information seeking. IEEE Software 11, 70–77 (1994)

Face Sketch Synthesis via Multivariate Output Regression

Liang Chang[1], Mingquan Zhou[1], Xiaoming Deng[2], Zhongke Wu[1], and Yanjun Han[3]

[1] College of Information Science and Technology, Beijing Normal University, Beijing 100875, P.R. China
[2] Institute of Software, Chinese Academy of Sciences, Beijing 100190, P.R. China
[3] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.R. China
changliang@bnu.edu.cn

Abstract. This paper presents a multivariate output regression based method to synthesize face sketches from photos. The training photos and sketches are divided into small image patches. For each pairs of photo patch and its corresponding sketch patch in training data, a local regression model is built by multivariate output regression methods such as kernel ridge regression and relevance vector machine (RVM). Compared with commonly used single-output regression, multivariate output regression can enforce the synthesized sketch patches with structure constraints. Experiments are given to show the validity and effectiveness of the approach.

Keywords: Face sketch synthesis, multivariate regression.

1 Introduction

Face sketch synthesis is useful in applications such as face recognition and digital entertainments. In applications such as law enforcement, when the face photo of a suspect is not available, traditional photo based face recognition can not be used. Then a database of face sketches is needed for retrieval and recognition. However, to synthesize a face sketch from a photo is challenging, because the sketch generation mechanism of an artist is difficult to be approximated by rules. Previous methods for face sketch synthesis, most of which are based on global principle component analysis (PCA) or local k-nearest neighbor, can hardly approximate the intrinsic nonlinear mapping between face photo and sketch.

Several studies have been conducted on sketch synthesis. Tang[5] developed an eigentransform based algorithm, in which the transformation between photos and sketches is assumed to be linear. However, the assumption is hard to be satisfied in practice. Liu[3] presented a method which is similar in the spirit to local linear embedding[4,7]. For each input photo patch, the nearest neighbors are found, and reconstruction weights are computed. Preserving the neighbor relationships and weights, the desired sketch patch can be obtained. This method needs a carefully chosen of the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 555–561, 2011.
© Springer-Verlag Berlin Heidelberg 2011

number of nearest neighbors. Wang[2] proposed a method using a multiscale Markov random field model, which learns the face structure across different scales. As reported in [2], this method is relatively time-consuming due to use an inference procedure with belief propagation.

In this paper, a novel face sketch synthesis method based on multivariate output regression is presented. Multivariate output regression models are popular in computer vision and machine learning. They are widely used in applications such as handwritten digits reconstruction and 3D human pose estimation[8]. In multivariate regression, the input-output pairs of the training set are both vectors.

In our method, the face photos and sketches are divided into small overlapped patches for training and synthesis. For each pair of patches, multivariate output regression model is built to learn the vector-valued function between them. The input of the regression function is the feature vector of the small photo patch, and the output is the feature vector of the corresponding sketch patch. Since the number of variables in the regression function is very large, the regression problem is often ill-posed. We use different regularization methods including the ridge regression [1] and relevance vector regression machine(RVM) [9] to solve the regression problem. In RVM, Gaussian priors are given for each parameter and many basis vectors are of the coefficient of zero, so the regression model has sparse property.

Our method has advantages in two aspects:

(1) Compared with the single-variable regression, in which the output is scalar value, the outputs in multivariate regression are vectors, which can enforce the texture structure constraints of face sketch and approximate the local texture of sketch patches well;
(2) Our method is nonlinear intrinsically. We use kernel bases to obtain the nonlinear mapping between local photo and sketch regions, and it can greatly enhance the sketch synthesis performance.

2 Face Sketch Synthesis Method

In this section, face sketch synthesis method via multivariable output kernel regression is proposed. The method is based on the regression of local image patches, and ridge regression and relevance vector machine are used.

2.1 Multivariate Output Regression

The multivariate regression model maps the input feature vector to an output vector. Suppose the output vector to be $\mathbf{y} \in R^m$ and the input vector to be $\mathbf{x} \in R^d$, the mapping between $\mathbf{x}$ and $\mathbf{y}$ can be approximated with a linear combination of basis functions $\{\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \cdots, \varphi_p(\mathbf{x})\}$ as follows

$$\mathbf{y} = \beta_1\varphi_1(\mathbf{x}) + \cdots + \beta_p\varphi_p(\mathbf{x}) \tag{1}$$

where $\{\beta_k\}_{k=1}^{n} \in R^m$ are weight vectors.

We transform $\{\beta_k\}_{k=1}^{n}$ to a weight matrix $\mathbf{B}_{m\times p}=(\beta_1,\beta_2,\cdots\beta_p)$ and the basis functions into a R^p valued function $f(\mathbf{x})=(\varphi_1(\mathbf{x}),\varphi_2(\mathbf{x}),\cdots\varphi_p(\mathbf{x}))^T$. Equation (1) can be formulated as

$$\mathbf{y}=\mathbf{B}f(\mathbf{x})+\varepsilon \tag{2}$$

where ε is residual error vector, and $f(\mathbf{x})$ is the feature vector of input vector $\mathbf{x}$.

For underdetermined system, the solution to equation (2) is not unique[1]. To seek for the unique solution to the weight matrix $\mathbf{B}$, minimization problem of the square residual error with a regularization item is required to be solved:

$$\min_{\mathbf{B}} \sum_{i=1}^{n} \left\|\mathbf{B}f(\mathbf{x}_i)-\mathbf{y}_i\right\|^2+R(\mathbf{B}) \tag{3}$$

where $R(\mathbf{B})$ is a regularization item on $\mathbf{B}$.

After simple manipulations, equation (3) can be transformed into the following matrix formulation:

$$\min_{\mathbf{B}} \left\|\mathbf{BF}-\mathbf{Y}\right\|^2+R(\mathbf{B}) \tag{4}$$

where $\mathbf{Y}=(\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_n)$ and $\mathbf{F}=(f(\mathbf{x}_1),f(\mathbf{x}_2),\cdots,f(\mathbf{x}_n))$.

We build the feature vector $f(\mathbf{x})$ by kernel bases as $f(\mathbf{x})=(K(\mathbf{x},\mathbf{x}_1),\cdots,K(\mathbf{x},\mathbf{x}_n))^T$, in which $K(\mathbf{x}_i,\mathbf{x}_j)$ is chosen as Gaussian kernel $K(\mathbf{x}_i,\mathbf{x}_j)=\exp(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2/\sigma^2)$ with σ estimated from scatter matrix of the training data. Therefore, $\mathbf{F}$ in equation (4) is taken as the kernel matrix of the training data, i.e. $\mathbf{F}=(K(\mathbf{x}_i,\mathbf{x}_j))$.

2.2 Ridge Regression and Relevance Vector Machine

In our problem, we aim to synthesize a face sketch from a face photo. The synthesis problem is transformed into a regression problem(See equation (2)), in which the input vector $\mathbf{x}$ is of a photo patch and the output vector $\mathbf{y}$ is the corresponding sketch patch of $\mathbf{x}$. However, the number of unknown variables in the weight matrix $\mathbf{B}$ is large, the regression is usually ill-posed, and conventional least square regression can no longer be used [1]. In order to get the unique solutions, regression methods with different regularization items are adopted. In this paper, the ridge regression and relevance vector machine are used.

(1) Ridge Regression

In ridge regression, the weight matrix $\mathbf{B}$ is estimated by minimizing the sum of square residual error and a regularization term[1]. The regularization term of the weight matrix takes the form of $R(\mathbf{B})=\lambda\left\|\mathbf{B}\right\|^2$, thus the ridge regression takes the form as follows:

$$\min_{\mathbf{B}} \|\mathbf{BF}-\mathbf{Y}\|^2 + \lambda \|\mathbf{B}\|^2 \quad (5)$$

where λ is the regularization parameter.

(2) Relevance Vector Machine

Relevance vector machines (RVM) are widely used in classification and regression in computer vision and machine learning [9]. It belongs to sparse Bayesian approaches. The column vector $\mathbf{b}_i$ of $\mathbf{B}$ takes the prior of the form $\|\mathbf{b}_i\|^{-\lambda}$. If λ is large, the regularization term makes many parameters in $\mathbf{B}$ to be zeros, thus the regression model is sparse. With the log likelihood priors, RVM takes the minimization form as follows:

$$\min_{\mathbf{B}} \|\mathbf{BF}-\mathbf{Y}\|^2 + \lambda \sum_k \log \|\mathbf{b}_k\| \quad (6)$$

where λ is the regularization parameter.

2.3 Face Sketch Synthesis Method

In this section, the face sketch synthesis method via multivariate regression is proposed. We divide the training face photos and sketches into a set of overlapped patches. For each pairs of training photo and sketch patch at the same location, the multivariable output kernel regression model is built and the weight matrix $\mathbf{B}$ is estimated (See Section 2.2). Finally, we can reconstruct the sketch image of a test face photo by the estimated weight matrix $\mathbf{B}$. We divide the test face photo into a set of overlapped patches, and denote $f(\mathbf{x}_i)$ to be the feature vector of a photo patch $\mathbf{x}_i$, and then the corresponding sketch patch can be reconstructed by $\mathbf{B}f(\mathbf{x}_i)$.

Algorithm: Face Sketch Synthesis via Multivariable Output Regression
Input: A training set of face photos and the corresponding sketches, and a test face photo.
Output: A synthesized sketch of the test face photo.

(1) All the training photos and sketches and the test photo are divided into overlapped patches.
(2) For each image patch $\mathbf{x}_i$ of the test photo

a) Build the ridge regression or relevance vector machine model with photo and sketch patch pairs of the same location in the training images, and then estimate the weight matrix $\mathbf{B}$ (See Section 2.2).
b) Reconstruct the sketch patch of the photo patch $\mathbf{x}_i$ by multiplying the weight matrix $\mathbf{B}$ and the feature vector $f(\mathbf{x}_i)$ of the photo patch $\mathbf{x}_i$, i.e. $\mathbf{B}f(\mathbf{x}_i)$.

(3) Reconstruct the sketch image by enforcing smoothness between the overlapped regions of the patches.

3 Experiments

The experiments are based on face photo-sketch database from CUHK student database [6]. Since the same face component is roughly in the same region of photos and sketches due to geometry alignment in the preprocessing step, we divide the training face photos and sketches into a set of regions by scanning the whole image area. We use 88 faces for training and 100 faces for testing. For each face photo, a sketch by an artist and a photo taken in the front pose are given. The feature vectors of the photos and sketches are represented with Gaussian kernel bases (See $f(\mathbf{x})$ in equation (2)) by the gray values of the corresponding photo and sketch patches. The size of all the face and sketch images are 160×120. The size of small patches 7 × 7. For adjacent patches, we keep half of the patch overlapped.

In Fig. 1 and Fig. 2, we demonstrate the experiment results by our method with ridge regression with Gaussian kernel bases, and the regularity parameter λ is set to be 1.0. In Fig. 1, we compare our results with those by eigen-face method [5] and nearest neighbor method[3]. It can be seen that our method can synthesize the hair region well, which is hard for the eigen-face method[5]. Moreover, our method can avoid the selection of neighborhood number in the nearest neighbor method[3]. Experiments show that the results of our method can approximate the sketches drawn by the artist in the style and structures.

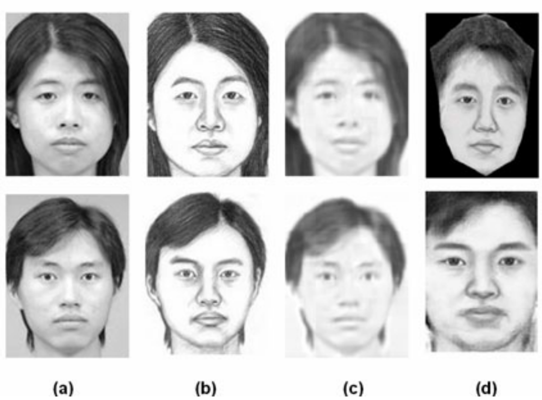


Fig. 1. (a) Photos; (b) sketches drawn by artists; (c) sketches by our method using kernel ridge regression; (d) sketches by the eigen-face method in [5](the first line) and k-nearest neighbor method in [3] (the second line), which are both from [2]

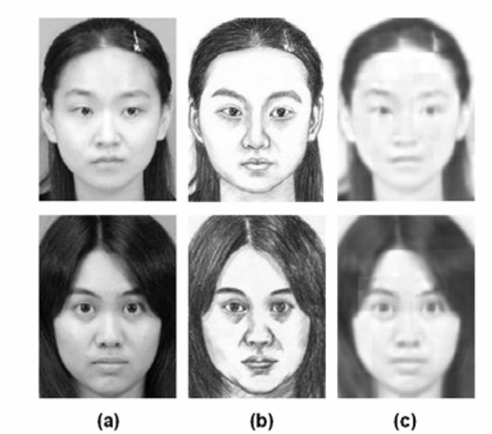


Fig. 2. (a) Photos; (b) sketches drawn by artists; (c) sketches by our method using kernel ridge regression

4 Conclusion

In this paper, we propose a face sketch synthesis method via multivariable output regression. In our method, all the face photos and sketches are divided into overlapped patches, and then multivariable output regression such as kernel ridge regression and relevance vector machine are used to build the regression model between the photo patch and sketch patch. Experiments show that the synthesized sketch can resemble the sketch draw by artists. Our research indicates that multivariable output regression approach is effective in face sketch synthesis.

Acknowledgments. The work is supported by National Natural Science Foundation of China under Grant No. 60736008/F010303 and 61005039, the open projects program of National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Sciences under Grant No. 20090096.

References

1. Bishop, C.: Neural Networks for Pattern Recognition. Oxford Univ. Press, Oxford (1995)
2. Wang, X., Tang, X.: Face Photo Sketch Synthesis and Recognition. IEEE Trans. on PAMI (2009)
3. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A Nonlinear Approach for Face Sketch Synthesis and Recognition. In: Proceedings of CVPR (2005)
4. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through Neighborhood Embedding. In: Proceedings of CVPR (2004)
5. Tang, X., Wang, X.: Face Sketch Recognition. IEEE Trans. on CSVT 14, 50–57 (2004)

6. The Chinese University of Hong Kong, Student sketch database (2009), http://www.mmlab.ie.cuhk.edu.hk/facesketch.html
7. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
8. Agarwal, A., Triggs, B.: Recovering 3D human pose from monocular images. IEEE Trans. on PAMI 28, 44–58 (2006)
9. Tipping, M.: The relevance vector machine. Neural Information Processing Systems (2000)

Experimental Studies of Visual Models in Automatic Image Annotation

Ping Guo, Tao Wan, and Jin Ma

Image Processing and Pattern Recognition Laboratory,
Beijing Normal University, Beijing 100875, China
pguo@ieee.org

Abstract. Semantic image annotation can be viewed as a mapping procedure from image features to semantic labels, by the steps of image feature extraction and image-semantic mapping. The features can be low-level visual features, such as color, texture, shape, etc., and the semantic labels can be related to the knowledge of human on the image understanding. However, these linear representations are insufficient to describe the complex natural scene. In this paper, we study currently existing visual models that are able to imitate the way the human visual system acts for the tasks of object recognition and scene interpretation. Therefore, it is expected to bring a better understanding to the image visual content in human cortex will. In the experiments, there are three state-of-the-art visual models are investigated for the application of automatic image annotation. The results demonstrate that with our proposed strategy, the annotation accuracy is improved comparing to the most used low-level linear representation features.

Keywords: Automatic image annotation; Semantics; Visual models; Human visual system; Low-level features.

1 Introduction

In computer vision research field, automatic semantic image annotation is attracted many researchers' interest. Semantic image annotation can be viewed as a mapping procedure from image features to semantic labels, by the steps of image feature extraction and image-semantic mapping. When the image annotation is posed as a classification problem, it becomes a mapping problem from low-level visual feature to high-level semantic label. Low-level visual features (color, shape, texture, edges) are easy to be dealt with computer, while high-level semantic labels are related to the knowledge of human on the image understanding.

In visual feature extraction, most studies use linear representations. The methods, such as Gabor transform, discrete cosine transform and wavelet transform, are adopted. But for a complex natural scene, perceptually distinct image regions produce response patterns that are highly overlapping and cannot be easily distinguished using low-level, linear representations [1]. As we know, human visual system outperforms the best machine vision system in any measurements. A fundamental function of the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 562–570, 2011.
© Springer-Verlag Berlin Heidelberg 2011

visual system is to encode the building blocks of natural scenes that subserve visual tasks such as object recognition and scene understanding. Therefore, it is expected to bring us better understanding to image visual feature through emulating object recognition in human cortex will. In this paper, we will experimental study currently existing visual models, and apply to image annotation problem.

In the experiments, three state-of-the-art visual models are investigated for automated image annotation application problem. These models include the model introduced by Serre *et al.* [2], model proposed by Mutch *et al.* [3], and model proposed by Karklin *et al.* [1], we call them Model-A, Model-B, and Model-C, respectively. The performances of three visual models are evaluated by using various natural images. In our proposed strategy, the neural population acts in the visual models are used as image features, and the support vector machine (SVM) classifier is trained for semantic image annotation.

1.1 Related Work

The recognition of different kinds of object categories with respect to illumination conditions, viewpoints from different positions, and diverse backgrounds is one of the major challenges for computer vision [4]. The experiments also show that human brains outperform the best machine vision system in similar cognitive and detective tasks [5]. Therefore, it is essential for researchers to understand how visual cortex recognizes objects as well as to emulate object recognition procedure.

In 1946, Gabor filter has been proposed to justify the sensory coding in the early vision studies [6] [7]. It has been proved to be a good model of cortical simple cell receptive fields, and the filtered results were used to classify images. Another approaches have focused on examining receptive fields of a more complex type [8], such as position or scale invariance, and constructing different models [5] [9] [10] [11], which learn to code the pixel intensities of a patch of texture or edge. However, much of work has concentrated on fitting mathematical models to image data or has been motivated by the specific computational goals [12]. For example, in a standard model of complex cells [10], the "energy" model including two localized and oriented features (typically 90° out of phase Gabor functions) is convolved with the image, and the outputs are squared and summed to give the neuron's response. In addition, the model described in [11] is to maximize the sparseness of locally pooled energies that correspond to complex cell outputs to show emergence of a topographic organization.

Another type of vision model stressed importance to a quantitative theory of the ventral stream of visual cortex [1] [13]. They extracted image features much like what the simple and complex cells in human brain would do, and then labeled images with the help of these features. Serre *et al.* [2] have improved a cognitive system that closely followed the organization of visual cortex and build an invariant feature representation by alternating between a template matching and a maximum pooling operation. They made the features complex to scale and position after each layer, and used SVM or boosting as the classifier. Mutch and Lowe [3] introduced an object recognition method based on sparse features with limited receptive fields. They made a few improvements on the Serre's model. In recent years, Karklin and Levicki proposed a new model using complex cell properties [1]. It generalized over similar images, higher-level visual neurons encode statistical variations that characterize local image regions. The subsequent section will give a brief introduction for these models.

1.2 Visual Model Overview

Serre *et al.* [2] introduced a general framework for the recognition of complex visual scenes, which is motivated by the findings in biology. The key element in the approach is a set of scale and position-tolerant feature detectors, which agree quantitatively with the tuning properties of cells along the ventral stream of visual cortex. These features are able to adaptive to the training set. They demonstrate the strength of the approach on a range of recognition tasks from invariant single object recognition in clutter to multiclass categorization problems and complex scene understanding tasks. They also showed a universal feature set learned from a set of natural image unrelated to any categorization task and achieved good performance.

Mutch and Lowe [3] developed an improved approached based on Serre's model by incorporating sparsity and localized intermediate-level features. They first applied Gabor filters at all positions and scales. Feature complexity and position/scale invariance are then built up by alternating template matching and max pooling operations. Several refinement steps are taken to increase the sparsity by constraining the number of feature inputs, lateral inhibition, and feature selection. The model is a partial implementation of the standard model of object recognition in cortex. The experiments have indicated that these modifications can improve classification performance.

Karklin and Lewicki have proposed a hierarchical model in [14, 15], which not only described sparse marginal densities and magnitude dependencies, but also captured a variety of joint density functions that are consistent with previous observations and theoretical conjectures. They also developed a new vision model based on this hierarchical model [15], which summarized the patterns of correlations for a given type of image using the covariance matrix of the data. In this model, every image patch corresponds to a latent variable, which encodes the image distribution consistent with the input image. An important advantage of this approach is that, it learns a general set of features that are determined by the statistical structures in natural images.

The paper is organized as follows. Section 2 describes the image annotation method using these three visual models. The experimental results and discussions are present in Section 3. The conclusions and feature work are summarized in Section 4.

2 Image Annotation with Visual Models

In this work, we experimental study these three state-of-the-art visual models applied to semantic image annotation problem. The basic idea is to adopt the visual model for image feature extraction, then feed these features to a SVM classifier to implement image annotation. The annotation procedure with these three visual models are described in details in the following subsections.

2.1 Image Annotation Using the Model-A

The detail annotation procedure with Model-A proposed by Serre *et al.* [2] is described as follows:

Step 1: Apply feature extraction method to all training and test images.

- S1 units: Apply a set of Gabor filters to each image. The filters are arranged to form a pyramid of scales, spanning a range of sizes from 7×7 to 37×37 pixels in step of two pixels. To keep the number of units tractable, four orientations (0°, 45°, 90°, and 135°) are considered, thus leading to 64 different filters in total (16 scales × 4 orientations).
- C1 units: C1 units pool over afferent S1 units from the previous layer to get the local maximum value with the same orientation and from the same scale band. The parameter setting can be found in [2]. Each scale band contains two adjacent filter sizes (there are eight scale bands for a total of 16 S1 filter sizes). For instance, scale band 1 contains S1 filters with sizes 7×7 and 9×9. The scale band index of the S1 units also determines the size of the S1 neighborhood NS×NS over which the C1 units pool. Again, this process is performed for each of the four orientations and each scale band independently.
- S2 units: In the S2 layer, units pool over afferent C1 units from a local spatial neighborhood across all four orientations. S2 units behave as radial basis function (RBF) units. Each S2 unit response depends in a Gaussian-like way on the Euclidean distance between a new input and a stored prototype. That is, for an image patch X from the previous C1 layer at a particular scale S, the response r of the corresponding S2 unit is given by:

$$R = \exp(-\beta \|X - P_i\|^2) \cdot \quad (1)$$

  In the experiment, we take β as 1/800. P_i is one of the N features (center of the RBF units) learned during training. To initialize the P_i in RBF neural network, a simple sampling process is applied that during training, a large pool of prototypes of various sizes and at random positions are extracted from a target set of images. These prototypes are extracted at the level of the C1 layer across all four orientations, i.e., a patch P_0 of size $n \times n$ contains $n \times n \times 4$ elements.
- C2 units: The element of the final feature vector is computed by taking a global maximum value over all scales and positions for each S2 type over the entire S2 lattice, i.e., the S2 measures the match between a stored prototype Pi and the input image at every position and scale; we only keep the value of the best match and discard the rest. As a result, a feature vector of $N \times 1$ dimension is formed for each image.

Step 2: Apply a SVM classifier with RBF kernel to the feature vectors above to get the separating surfaces, then assign semantic label to corresponding image.

2.2 Image Annotation Using the Model-B

The Model-B [3] is implemented and summarized as follows:

Step 1: For the image layer, we convert the image to grayscale and scale the shorter edge to 140 pixels while maintaining the aspect ratio, and then we create an image pyramid of 10 scales, each a factor of $2^{1/4}$ smaller than the last.

Step 2: For the Gabor filter (S1) layer, the S1 layer is computed from the image layer by centering 2D Gabor filters with a full range of orientations at each possible position and scale. The base model follows [2] and uses 4 orientations. The Gabor filters are 11×11 in size, and are described by

$$G(x,y) = \exp(-(\mathbf{X}^2 + \gamma^2\mathbf{Y}^2)/(2\times\delta^2))\times\cos(2\pi\mathbf{X}/\lambda), \tag{2}$$

where $\mathbf{X} = x\cos\theta - y\sin\theta$ and $\mathbf{Y} = x\sin\theta + y\cos\theta$. x and y vary between -5 and 5, and θ varies between 0 and π. The parameters γ (aspect ratio), δ (effective width), and λ (wavelength) are all taken from [2] and are set to 0.3, 4.5, and 5.6 respectively.

Step 3: Local invariance (C1) layer pools nearby S1 units to create position and scale invariance over larger local regions, and as a result can also subsample S1 to reduce the number of units. For each orientation, the S1 pyramid is convolved with a 3D max filter, 10×10 units across in position and 2 units deep in scale. A C1 unit's value is simply the value of the maximum S1 unit (of that orientation) that falls within the max filter. Due to the pyramidal structure of S1, we are able to use the same size filter for all scales. The resulting C1 layer is smaller in spatial extent and has the same number of feature types (orientations) as S1.

Step 4: In intermediate feature (S2) layer, at every position and scale in the C1 layer, we perform template matches between the patch of C1 units centered at that position/scale and each of d prototype patches. These prototype patches represent the intermediate-level features of the model.

Step 5: In global invariance (C2) layer, we create a d-dimensional vector that is classified using the linear SVM classifier.

2.3 Image Annotation Using the Model-C

In Model-C, Karklin assumed that the individual image patch $\mathbf{x}$ is with multivariate Gaussian probability distributions [1],

$$p(\mathbf{x}|\mathbf{y}) = N(0,\mathbf{C}), \tag{3}$$

where covariance matrix $\mathbf{C}$ is represented by a set of basis functions $\mathbf{A}_j$,

$$\log\mathbf{C} = \sum y_j\mathbf{A}_j, \tag{4}$$

where $\mathbf{A}_j$ is a symmetric matrix of the same size with the covariance matrix $\mathbf{C}$, and can be illustrated as:

$$\mathbf{A}_j = \sum_k w_{jk}\mathbf{b}_{jk}\mathbf{b}_{jk}^T. \tag{5}$$

Every patch has a latent variable y that has a different set of weights w_{jk}, corresponding to an expansion or contraction along vector $\mathbf{b}_{jk}$. Different image patches correspond to different latent variables y_j, which is regarded as neural population acts. In the experiment, we take y_j as the feature vectors.

There are three main steps to build the Model-C [1]:

Step 1: Image pre-processing is performed to ensure that all the image patches are with Gaussian distribution. To emulate the transformation at the retinal cone cells [14] [15], the original images are transformed to grayscale. Then after pixel intensities log10-transformed, the images are filtered with a Gaussian low-pass filter. Then, the images are down-sampled with the rate of 2:1. Each image in the database is divided into non-overlapping 20×20 image patches. Every patch is transferred into a column vector. The desirable patches are selected as the training data to form a large matrix. The mean luminance value is calculated for the large matrix and subtracted from each patch. This is "whitened" [16] all image patches to remove global correlations and to normalize the variance.

All image patches are formulated in a matrix called ***Pool***, which is a 400×N matrix (N is the number of image patches). Let ***mPool*** be the mean luminance value of ***Pool***, then the covariance **C** is calculated as,

$$\mathbf{C} = \boldsymbol{rePool} * \boldsymbol{rePool'}, \tag{6}$$

where ***rePool*** is the matrix ***Pool*** minus its mean value, and

$$\boldsymbol{rePool} = \boldsymbol{Pool} - \boldsymbol{mPool}. \tag{7}$$

Then the eigenvalues ***EPool*** and eigenvectors ***DPool*** are computed using the covariance matrix. The whitening is done by this,

$$\boldsymbol{whitex} = 1/\mathrm{sqrt}(\boldsymbol{DPool}) \times \mathbf{C}^{-1/2} \times \mathbf{x}, \tag{8}$$

where **x** is a column vector in ***rePool***, and ***whitex*** represents the whitened patch.

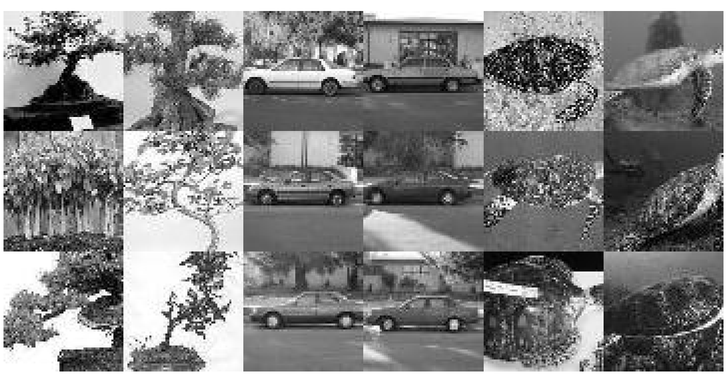


Fig. 1. Sample images used in the experiment

Step 2: In the feature extraction stage, the visual model parameters are estimated. After the training process, the latent variable y is computed for every image patch based on the perceptual model in the dataset. These latent variables of image patches are regarded as neuron population acts to objects in image, we take them as feature vectors. The part of feature vectors are used as training data for the SVM classifier and others are the testing data.

Step 3: The SVM classifier is used to perform the annotation task on images.

3 Experimental Results and Discussions

3.1 Summary of the Experiments

The evaluation of an image annotation system requires three components: an image database with manually produced annotations, which is used as training and verification dataset, a strategy to construct annotation system, and a set of measures to verify annotation performance. In the experiment, the image database is consisted of 351 images partly selected from the Caltech 101 object categories. There are 3 classes in the semantic label set, including “Bonsai”, “Car side”, and “Hksbil”. Some images used in the experiments are shown in Fig. 1.

In the experiment, each image is assigned with a caption of only one label. In the Model-A and Model-B, the Gabor filters are applied to each image and it is decomposed into a set of regions. The filters are arranged to form a pyramid of scales, spanning a range of sizes from 7×7 to 37×37 pixels in steps of 2 pixels. To keep the number of units tractable, we consider four orientations (0°, 45°, 90°, and 135°), thus leading to 64 different filters total (16 scales × 4 orientations).

In the Model-C, pixel intensities are log-transformed, corresponding roughly to the transformation at the retinal cone cells, and the images are low-pass filtered. Taking the same experiment setup as described in [15], 20×20 pixel image patches are extracted from the entire dataset without using Gabor filter. The mean luminance value is subtracted from each patch, thus speeding up the model training process without significant influence on the performance. All the image patches are whitened in order to remove global correlations and normalize the variance. This allows the model to encode only the deviations of each image distribution from the global statistics. After the pre-processing approach, the neuron population acts y are calculated by training the model with image patches. In training stage, the values of y are estimated with EM algorithm. With obtained neuron coding y as the input of the SVM classifier, image is annotated using class label of SVM output. In this step, about half of samples are taken as the training samples and the remains as test samples. Finally, the annotation precision of every image concept in the test set is computed as:

$$\text{precision} = w_c \,/\, w_{auto}, \tag{9}$$

where w_{auto} is obtained from the annotation system, and w_c is the SVM classifier output.

3.2. Experimental Results and Discussions

Three visual models have been developed and evaluated on a variety of natural images. In Table 1, we can see that for different classes these three visual models produced various results. For example, in some classes such as "Car side", "Hawksbill", some visual models can achieve good results, however, in some other classes like "Bonsai", the performance is not satisfied. It shows that the performance of the visual models is affected by the factors of the complexity of the shape and texture of the target object.

In the "Bonsai" class, the Model-A obtains the better precision rate than the Model-B and Model-C. While for "Car side" class, the Model-B gets the highest annotation precision. From the experiments, it can be found that the Model-B tends to produce more stable results in comparison with the Model-A and the Model-C. The reason lies in that the Model-B uses the Gabor filter with the multiple orientations comparing to the one used in the Model-A with four orientations. There are some additional improvements in the Model-B to further enhance the performance.

It is worth noticing that the Model-C has the highest precision values than the other two models for "Hawksbill" class. This illustrates that Model-C is suitable to encoding those image with rich texture information. Original, the main purpose of the Model-C is to describe the natural scene images with texture features. For those image, such as the in-door images, Model-A is more suitable to encoding the features compare to Model-C.

In general, from experiments it is found that annotation accuracy is improved comparing to the most used low-level linear representation features [17].

Table 1. Annotation precision using three visual models

Concept	Model-A	Model-B	Model-C
Bonsai	57.5	51.0	55.8
Car side	88.0	94.6	58.2
Hawksbill	70.0	52.8	85.0

4 Conclusions and Feature Work

In this paper, we explore three novel visual models that are able to imitate the way of the high-level human perception to interpret the image contents. These three models have been applied to the image annotation problem and show the superior results in comparison with the conventional low-level linear representation features. The proposed methodology could serve as a critical part in a content-based image retrieval system to enhance the search query performance in terms of human judgment. In future, we would like to study most these three visual models with respect to visual descriptors. Also we are interested in investigating the extension of our work to the semantic image retrieval system. Experimental results demonstrate that with our proposed strategy, the annotation accuracy is improved comparing to the most used low-level linear representation features.

Acknowledgments. The research work described in this paper was fully supported by the grants from the National Natural Science Foundation of China (Project No. 90820010, 60911130513). The authors would like to express their thanks for Mr. Rukun Hu for his work in parts of experiment.

References

1. Karklin, Y., Levicki, M.S.: Emergence of Complex Cell Properties by Learning to Generalize in Natural Scenes. Nature 457, 83–86 (2009)
2. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust Object Recognition with Cortex-Like Mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 411–426 (2007)
3. Mutch, J., Lowe, D.G.: Object Class Recognition and Localization Using Sparse Features with Limited Receptive Fields. International Journal of Computer Vision 80, 45–57 (2008)
4. Gabor, D.: Theory of Communication. J. IEE 93, 429–459 (1946)
5. Field, D.: What is the goal of sensory coding? Neural Computation 6, 559–601 (1994)
6. Hubel, D.H., Wiesel, T.N.: Receptive Fields, Binocular Interaction and Functional Architecture in The Cat's Visual Cortex. J. Physiology 160, 106–154 (1962)
7. Cavanaugh, J.R., Bair, W., Movshon, J.A.: Nature and Interaction of Signals from The Receptive Field Center and Surround in Macaque V1 Neurons. J. Neurophysiology 88, 2530–2546 (2002)
8. Heeger, D.J., Simoncelli, E.P., Movshon, J.A.: Computational Methods of Cortical Visual Processing. Proc. Natl Acad. Sci., 623–627 (1996)
9. Hyvarinen, A., Hoyer, P.: A Two-layer Sparse Coding Model Learns Simple and Complex Cell Receptive Fields and Topography from Natural Images. Vision Res. 42, 241–2423 (2001)
10. Hubel, D., Wiesel, T.: Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas of the Cat. J. Neurophysiology 28, 22–289 (1965)
11. Bruce, C., Desimone, R., Gross, C.: Visual Properties of Neurons in a Polysensory Area in the Superior Temporal Sulcus of the Macaque. J. Neurophysiology 46, 369–384 (1981)
12. Comon, P.: Independent Component Analysis, a New Concept? Signal Processing 36, 287–314 (1994)
13. Ullman, S., Vidal-Naquet, M., Sali, E.: Visual Features of Intermediate Complexity and Their Use in Classification. Nature Neuroscience 5, 68–687 (2002)
14. Karklin, Y., Levicki, M.S.: A Hierarchical Bayesian Model for Learning Nonlinear Statistical Regularities in Nonstationary Natural Signals. Neural Computation 17, 397–423 (2005)
15. Karklin, Y.: Hierarchical Statistical Models of Computation in the Visual Cortex. PhD thesis, Carnegie Mellon University (2007)
16. Perlibakas, V.: Distance Measures for PCA-based Face Recognition. Pattern Recognition Letters 25, 711–724 (2004)
17. Hu, R.K., Shao, S., Guo, P.: Investigating Visual Feature Extraction Methods for Image Annotation. In: Proceedings of 2009 IEEE International Conference on Systems, Man, and Cybernetics, pp. 3211–3216 (2009)

An Image Segmentation Method for Chinese Paintings by Combining Deformable Models with Graph Cuts

Ning He[1] and Ke Lu[2]

[1] School of Information, Beijing Union University,
Beijing, 100101, China
xxthening@buu.edu.cn
[2] College of Computing & Communication Engineering,
Graduate University of Chinese Academy of Sciences,
Beijing, 100049, China
luk@gucas.ac.cn

Abstract. In recent years researchers have developed many graph theory based algorithms for image setmentation. However, previous approaches usually require trimaps as input, or consume intolerably long time to get the final results, and most of them just consider the color information. In this paper we proposed a fast object extraction method. First it combines deformable models information with explicit edge information in a graph cuts optimization framework. we segment the input image roughly into two regions: foreground and background. After that, we estimate the opacity values for the pixels nearby the foreground/background border using belief propagation (BP). Third, we introduce the texture information by building TCP images' co-occurrence matrices. Experiments show that our method is efficient especially for TCP images.

Keywords: Graph cuts, Deformable Model, Traditional Chinese Painting (TCP), Belief Propagation (BP), Po-Occurrence Matrix.

1 Introduction

Segmentation of Traditional Chinese Painting (TCP) images is a good first step to separate them from generalized images. With the steady growth of computer power, rapidly declining cost of storage, and ever-increasing access to the Internet, digital acquisition of information has become increasingly popular in recent years[1]. Many organizations have a large digital images content available for on-line access. Various museums are constructing digital archives of art paintings and preserve the original artifacts. More and more artists attempt to exhibit and sell their productions on the Internet. Effective indexing, browsing and retrieving art images is an important and imperative problem need to be addressed, while image segmentation is a crucial step.

Deformable models and Graph cuts are two of the most important frameworks that have been used in the last decade to solve the image segmentation problem. Recent studies such as [2-4] are in the favor of combining deformable model approaches and

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 571–579, 2011.
© Springer-Verlag Berlin Heidelberg 2011

graph cuts optimization to benefit from the advantages of deformable models in describing a wide variety of segmentation techniques and meanwhile take advantage of the fast global optimization associated with the use of graph cuts. The energy formulation of the model relaxes the global piecewise constant constraint in [3] and [5]. The energy will be optimized using graph cuts. In this paper, we introduces a new method for interactive segmentation that providing better boundary placement than deformable models segmentation and stronger region connectivity and less short-cutting than graph cuts methods. The experiment result shows that the technology is an effective method for the segmentation of TCP images.

2 The Features of TCP Images

Throughout its long history, TCP images have carried their own particular style. they look different from general image. Color and texture features are used here to characterisze the particularity of TCP images.

The color feature representation the corresponding percentage of occurrence of each color within a certain neighbourhood. Howe et al. have proposed in [6] a color-shape based method in which a quantized color image I' is obtained from the original image I by quantizing pixel colors in the original image. The authors tested on RGB, YUV, YIQ color space using SVM classifier. The result is in Table 1.

Autocorrelation [7] measures the coarseness of an image by evaluating the linear spatial relationships between texture primitives. Large primitives give rise to coarse texture and small primitives give rise to fine texture. If the primitives are large, the autocorrelation function decreases slowly with increasing distance whereas it decreases rapidly if texture consists of small primitives. Typically, TCP images have larger feature values compared to non-TCP images.

Table 1. Test result on color space

Classification rate on TCP test images			False classification rate on non-TCP test images		
RGB	YUV	YIQ	RGB	YUV	YIQ
0.902	0.942	0.940	0.117	0.09	0.076

3 The Proposed Method

Graph cuts algorithm[8,9] has established the relationship between the optimize of the energy function and the maximum flow/minimum cut algorithm in graph theory. In many cases can get the local optimal solution and in some cases get the global optimal solution. However, for a large amount of data processing, the graph cuts algorithm has the limitation in the memory requirements and computation velocity.

Different to reference [10], we consider each layer by accelerating the convergence of graph segmentation algorithm to reduce computation time. In this paper we use Gaussian Mixture Model (GMM), mean value and covariance matrix as the initial conditions, without having to use narrow-band. Although the number of nodes in every layer has not decreased, the parameters of images be obtained easily, this method can converge quickly and get the satisfied results.

Similar to reference [11], we use Gaussian Mixture Model (GMM) as the image color model. z is an array of image pixels, α and k are the non-transparent and the Gaussian component index arrays, respectively. Here α can only take 0 (background) or 1 (foreground). The parameters of each Gaussian component is defined as $\theta = \{\pi(\alpha,k), \mu(\alpha,k), \sum(\alpha,k)\}$, they are the weight of Gaussian component, mean vector and covariance matrix, respectively. The problem of image segmentation can be transformed into an deformable model optimization problem, and then use the graph guts method to solve. Here, the definition of the Gibbs energy function as:

$$\mathrm{E}(\alpha, \mathrm{k}) = \sum_i \mathrm{D}(\alpha_i, k) + \sum_{m,n} \mathrm{V}(\alpha_m, \alpha_n) \cdot \tag{1}$$

here, the optimal solution of α and k get the minimum of the energy function. Where $D(\alpha_{i,k}) = -\log P(\alpha_{i,k})$, $V(\alpha_m, \alpha_n) = [\alpha_m \neq \alpha_n]\exp(-\beta \|C_m - C_n\|^2)$, $[\alpha_m \neq \alpha_n]$ is the indicator function. The value is 0 when they are equal, otherwise is 1. $P(\alpha_{i,k})$ is Gaussian probability, $V(\alpha_m, \alpha_n)$ is smooth term.

4 Boundary Extraction

Belief propagation (BP) proposed by Pearl [12], are widely used in the encoding and decoding, stereo vision, image restoration and other fields. For acyclic graphs, belief propagation can get the global optimal solution, and for cyclic graph ,can only get the approximate solution. In practice, even for Markov Random Field (MRF) situations with the complex acyclic images can be obtained the optimal solution.

Consider using the MRF framework and the belief propagation algorithm to solving the boundary extraction problem. Each pixel in the image area R is as a MRF node. And the 4-connected neighboring pixels in R of boundary are included in the MRF. The aim of the image segmentation for the initial results in the foreground/background boundary width w to estimate the opacity. By minimizing the following defined energy function to get boundary of images:

$$E = E_{\text{Data}}(\alpha_i) + E_{\text{Smooth}}(\alpha_i, \alpha_j) + E_{\text{Texture}}(\alpha_{\text{i}}) \quad (i, j \in R) \tag{2}$$

The equation include in the data, smoothness and texture terms. Their significance are introduced as follow.

4.1 Data Term and Smoothness Term

The values $\alpha \in [0,1]$ is continuous, we use belief propagation algorithm to discrete it into K layers (e.g. $K = 10$), makes the each node in MRF can only be one of the K states.

To obtain estimates of the image pixel value α is equivalent to giving a label of each node.

The boundary of the foreground and background are smoothness, the adjacent pixels should have similar labels. If the value α of two nodes corresponding to the MRF in a clique has large differences, then the smooth item has greater penalty item, based on the above considerations, smooth item is defined as follows:

$$\mathrm{Smooth}(\alpha_i,\alpha_j)=1-\exp(-\beta\frac{(\alpha_i-\alpha_j)^2}{2\delta_s^{\,2}}) \tag{3}$$

where δ_ε is the experience value, i.e. it can be set to $K/2$, $\beta=\left(2\left\langle(x_p-x_q)^2\right\rangle\right)^{-1}$ is parameter like a regularized factor .

Smoothness term defined reflects the degree of similarity between adjacent pixels, and data item of the model is the fitting item of foreground and background. If a pixel is assigned a label closer to the real value, obviously the data will correspond to smaller penalty value, so that the total energy function is reduced. We define the following data item:

$$\mathrm{Data}\,(\alpha_i)=1-\frac{\mathrm{L_i(p)}}{\sum_{k=1}^{K}\mathrm{L_k(p)}} \tag{4}$$

To calculate the similarity $L_k(p)$ of pixels p and labels k. First, find N pixels, which is the smallest Euclidean distance with pixel p in foreground or background pixels were recorded as a collection of s_{obj} and s_{bkg}. Get pixel $p_i\in s_{\mathrm{obj}}$, note RGB color vector C_i; get pixel $p_j\in s_{\mathrm{bkg}}$, note the RGB color vector C_j. When mix C_i and C_j to α_k, the distance between pixel p and C is:

$$\left\|\alpha_k C_i+(1-\alpha_k)C_j-C\right\|^2 \tag{5}$$

where $\|\cdot\|$ is the 2- norm. Thus, the similarity of pixel p and labels $L_k(p)$ can be defined as follows:

$$L_{\mathrm{k}}(p)=\mathrm{Min}\ \exp(-\frac{D_{ij}(\alpha_k)}{\delta_d^{\,2}}) \tag{6}$$

The value of δ_d set by the user manually, generally it sets 15.

4.2 Texture Term

Co-occurrence matrix is the description of texture information in image areas , it has the advantages such as computation simply, can reflect the texture direction and distance. Some images have rich texture information relatively, contributed to the pixel to find the optimal label.

Co-occurrence matrix determined by the offset distance and offset angle. For convenience, we construct the 8-neighborhood of four co-occurrence matrix, corresponding to the 0°, 45°, 90°, 135° four cases, shown in Fig.1.

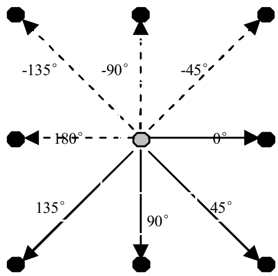


Fig. 1. Symbiotic matrix

For the pixels p of 8-connected must be considered the adjacent pair-pixels in 8 directions $\{p, p_i\}(i = 1...8)$. Let p and p_i are quantized a and b, corresponding to the direction of the foreground and background of the co-occurrence matrix were M_{obj} and M_{bkg}. If the angle formed by p and p_i is one of the solid arrow lines in Fig.1, the corresponding textures penalty value of foreground and background can be defined as follows, respectively:

$$\text{Penalty}_{obj} = \frac{M_{bkg}(a,b)}{M_{bkg}(a,b) + M_{obj}(a,b)}$$
$$\text{Penalty}_{bkg} = \frac{M_{obj}(a,b)}{M_{bkg}(a,b) + M_{obj}(a,b)}. \quad (7)$$

If the angle between p and p_i shown in Fig.2 is the dotted arrow line, obtained the value by symmetry. When the labels of p and p_i is α_i and α_j, respectively. To introduce the increment item in energy function.

$$\Delta E = (\alpha_i + \alpha_j) * \text{Penalty}_{obj} + [(1-\alpha_i) + (1-\alpha_j)] * \text{Penalty}_{bkg} \quad (8)$$

As MRF is 4-connected, co-occurrence matrix calculated using the eight different directions. In order to resolve inconsistencies between them, consider the following skills:

$$\Delta E = \Delta E(\alpha_i) + \Delta E(\alpha_j) = [\alpha_i * \text{Penalty}_{obj} + (1-\alpha_i) * \text{Penalty}_{bkg}] + [\alpha_j * \text{Penalty}_{obj} + (1-\alpha_j) * \text{Penalty}_{bkg}] \quad (9)$$

Therefore, $\Delta E(\alpha_i)$ and $\Delta E(\alpha_j)$ same as the data item, only consider the current pixel label information without considering the 8-connected neighboring pixel labels. so that the co-occurrence matrix texture information can be propagated in the 4 -connected MRF. In summary, texture items in energy function can be defined as follows:

$$\text{Texture}(\alpha_i) = \sum_{k=1}^{8} [\alpha_i * \text{Penalty}^{k}_{obj} + (1-\alpha_i) * \text{Penalty}^{k}_{bkg}] \cdot \quad (10)$$

Where k is the eight different directions.

5 Experimental Results

5.1 Running Time Analysis

We selected four representative TCP images from Internet. We implement the max flow-min cut to solve graph cuts method. Without using sophisticated data structures. The algorithm achieves the running time of $O(nm + n^2 \log U)$, where n is the number of nodes, m is the number of edges, and U is the largest edge weight. However, the simple topology of the graph used in this work makes the algorithm run much faster in practice. By obtaining the best polynomial fit to observed data, the minimum cut algorithm used in this work has running time of $O(n^{1.2})$. In Fig.2, the running time is shown as a function of the number n of nodes.

5.2 Comparison of Segmentation Accuracy

Fig.3 shows a sample comparison of performing the segmentation of graph cuts and the proposed method. Our approach clearly extracts the objects, while the graph cuts method failed to correctly identify some of the foreground pixels. Some background pixels wrongly identified as objects. The algorithm was implemented in Matlab 7.1 on a 2GHZ CPU with 1GB RAM. For quantitative assessment, we compared our segmentation results with the graph cuts by calculating the error rate. The results show in Table.2. Our method achieves better result. Fig.4 is in the same experimental conditions, we segmented the flower and birds respectively. The results are robust.

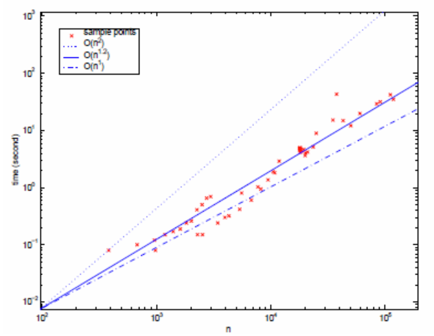


Fig. 2. The observed running time of the minimum cut algorithm used in our approach is $O(n^{1.2})$. The performance levels of $O(n)$ and $O(n^2)$ as shown, respectively.

(a) Original images (b) Results of our method (c) Results of graph cuts

Fig. 3. Compare results of proposed method and graph cuts method

(a) Original images (b) object is the flower (c) object is the birds

Fig. 4. The different segmentation results according to the different foreground

Table 2. Error Rate Comparation

Images	Proposed Method	Graph Cut Method
Person1	0.56%	1.82%
Tiger	0.35%	3.56%
Flower1	0.89%	1.29%

6 Conclusion

We have presented a graph cuts based deformation model approach to TCP images segmentation. First, we transform a multi-labels cut problem into a singe s-t minimum cut problem. Then, the method integrate the color and texture features of TCP images to extract the boundary. In the actual operation received good results.

Acknowledgments. This work was supported by the NSFC (China) Grant No.s 61070120, 60982145; Beijing national science fundation Grant No.4112021; Beijing educational commission science fundation; the opening project of Shanghai key laboratory of integrate administration technologies for nformation security, No.AGK2010005; the opening project of key laboratory of communication and information system (Beijing jiaotong university).

References

1. Jiang, S., Gao, W., Wang, W.: Classifying Traditional Chinese Painting Images. ICICS-PCM 2003, Singapore, December 15-18 (2003)
2. Criminisi, A., Sharp, T., Rother, C., Perez, P.: Geodesic Image And Video Editing. ACM Transactions on Graphics (2010)
3. Ding, L., Yilmaz, A.: Interactive Image Segmentation Using Probabilistic Hypergraphs. Pattern Recognition 43(5), 1863–11873 (2010)
4. Xu, N., Bansal, R., Ahuja, N.: Object Segmentation Using Graph Cuts Based Active Contours. In: CVPR, Los Alamitos, CA, USA, vol. 2, pp. 46–53 (2003)

5. Sun, J., Li, Y., Kang, S.B., Shum, H.-Y.: Flash Matting. ACM Trans. Graph 25(3), 772–778 (2006)
6. Liu, P., Jia, K., Wang, Z., Lv, Z.: A New and Effective Image Retrieval Method Based on Combined Features. In: Proc. IEEE Int. Conf. on Image and Graphics, pp. 786–790 (2007)
7. Oraintara, S., Nguyen, T.T.: Using Phase and Magnitude Information of the Complex directional Filter Bank for Texture Image Retrieval. In: Proc. IEEE Int. Conf. on Image Processing, vol. 4, pp. 61–64 (2007)
8. Boykov, Y., Jolly, M.P.: Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. In: International Conference on Computer Vision (ICCV), vol. I, pp. 105–112 (2001)
9. Kolmogorov, V., Zabih, R.: What Energy Functions Can Be Minimizedvia Graph Cuts? IEEE TPAMI 26, 147–159 (2004)
10. Lombaert, H., Sun, Y., Grady, L., Xu, C.: A Multilevel Banded Graph Cuts Method for Fast Image Segmentation. In: Proceedings of International Conference on Computer Vision, Beijing, China, pp. 1259–265 (2005)
11. McGuinness, K., O'Connor, N.E.: A comparative evaluation of interactive segmentation algorithms. Pattern Recognition 43, 434–444 (2010)
12. Felzenszwalb, P.F., Huttenlocher, D.R.: Efficient Belief Propagation for Early Vision. In: CVPR, IEEE Int. Conf. on Computer society, vol. 1, pp. 261–268 (2004)

Image Skeletonization Based on Curve Skeleton Extraction

Xiaoyan Hu, Bo Sun, Huiqin Zhao, Bin Xie, and Hao Wu

College of Information Science and Technology, Beijing Normal University
{huxy,tosunbo}@bnu.edu.cn

Abstract. Skeletonization is a transformation of an object in a digital image into a simplified representation of the original object. The skeleton of an image object is an abstraction of the original object which largely preserves the extent and connectivity of the original region while throwing away most of the boundary and interior pixels. In this paper, we propose a new method to calculate skeleton from 3D space instead of image space which has only two dimensions. Our method start with a contour of an object in an image, then inflate this two dimensional shape to a three dimensional mesh, and then apply a 3D mesh curve skeleton extraction algorithm to this intermediate three dimension mesh model. Finally, we project the resulting 3D curve skeleton back to image space and get the skeleton of the original shape or object in the image. Our method is noise insensitive. A little perturbation on shape would not change the structure of the resulting skeleton. Our method is relatively fast because it only generates a geometry mesh in contrast to compute a Voronoi graph. Our method preserves the topology as well as the shape.

1 Introduction

Skeletonization is a transformation of an object in a digital image into a simplified representation of the original object. The skeleton of an image object is an abstraction of the original object which largely preserves the extent and connectivity of the original region while throwing away most of the boundary and interior pixels. The notion skeleton was introduced by H. Blum as a result of the Medial Axis Transform (MAT) or Symmetry Axis Transform (SAT) [13]. The MAT determines the closest boundary point(s) for each point in an object. An inner point belongs to the skeleton if it has at least two closest boundary points. Skeletonization makes the object image easier to be read and use in many image processing and analysis applications such as feature analysis and pattern recognition and classification.

There are three major categories of skeletonization techniques to produce a skeleton representation of a certain object or region in an image. 1. Skeleton can be extracted based on distance transforms by detecting ridges in distance map of the boundary points. 2. Calculating the Voronoi diagram generated by the boundary points. 3. Using morphological thinning techniques based on erosion operations. Although these methods can compute skeleton for a certain object in an image, they are very sensitive to noise and small disturbance on boundary.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 580–587, 2011.
© Springer-Verlag Berlin Heidelberg 2011

In this paper, we propose a new method to calculate skeleton from 3D space instead of image space which has only two dimensions. Our method started with a contour of an object in an image, then we inflate this two dimensional shape to a three dimensional mesh, and then apply a 3D mesh curve skeleton extraction algorithm to this intermediate three dimension mesh model. Finally, we project the resulting 3D curve skeleton back to image space and we get the skeleton of the original shape or object in the image. The 3D curve skeleton extraction algorithm we used in this paper is mesh contraction method. This method is proposed by Oscar et al. [5], and can extract curve skeleton efficiently from a 3D mesh. One advantage of "mesh contraction" method is that it is noise insensitive. Another advantage of this curve skeleton extraction algorithm is that it is very fast. When applying this algorithm on a mesh, it would take only several iterations (about ten iterations) to convergence to a 3D curve skeleton which make the skeletonization interactive process. By using mesh contraction algorithm, our skeletonization method inherits these advantages. 1. Our method is noise insensitive. A little perturbation on shape would not change the structure of the resulting skeleton. 2. Our method is relatively fast because it only generates a geometry mesh in contrast to compute a Voronoi graph. 3. Our method preserves the topology as well as the shape.

2 Related Work

In this section, we discuss relate work in two fields which are relative to our method. One is traditional image skeletonizing methods and the other is curve skeleton extraction of objects in 3D space.

In image processing and computer vision, the skeleton of a region is proposed by Blum [13] and is defined via medial axis transformation (MAT) which is also known as grass fire transformation. There are three major categories of skeletonization techniques to produce a skeleton representation of a certain object or region in an image. 1. Skeleton can be extracted based on distance transforms by detecting ridges in distance map of the boundary points. [9, 22, 10, 12, 13] compute skeleton using wave front propagation or distance transform。 2. Another way is calculating the Voronoi diagram generated by the boundary points [16, 23]. 3. Using morphological thinning techniques based on erosion operations. [3, 24, 4].

Although these methods can compute skeleton for a certain object in an image, they are very sensitive to noise and small disturbance on boundary and the resulting skeleton can drastically different in its structure. This becomes a serious problem in digital image process. Several methods have been proposed to pruning “false” branches that are believed to be caused by noise in the outline. [16, 22, 10] Liu and Geiger et al. [26, 11] proposed a method based on self-similarity of a smooth outline curve to stabilize the skeleton extraction. Tang et al. [25] extracts skeleton from ribbon like shapes based on a new wavelet function.

In digital geometry processing, the extraction of curve skeletons from 3D models is a fundamental problem and is still a research challenging to find a simple and robust method that compute curve skeletons for 3D geometry mesh [7, 5]. Methods to extract curve skeleton for 3D objects can be roughly classified into two main categories, based on the underlying representation of 3D objects: volumetric methods and geometric methods.

Many exist methods work on volumetric discrete representation such as voxelized models [17, 21, 27] or distance field function [14, 28, 6, 18].

Geometric method work directly on polygon meshes or point sets. There two main categories methods, one obtain an approximate medial surface Voronoi diagram and prune it to get a curve-skeleton.[1, 8, 19], The other make use of Reeb graph based techniques via various real value functions defined on model surface [20, 15, 2]. Recently, Oscar et al. developed a simple method based on Laplacian smoothing to robustly extract curve skeleton [5]. Our method is mainly based on their work. We refer the reader to the comprehensive survey of [7] for more detailed discussion on curve skeleton extraction for 3D models.

3 Overview

The first step of our method is to transform 2D shape into a 3D mesh. In this stage, we apply snake algorithm to generate contour of object in source image. After we got contour, a two dimensional mesh will be generated within this contour. One simple way to generate this 2D mesh is to generate a quad grid per pixel within the contour and tessellate each quad grid into two triangles. This step results in a triangle mesh in image space. The next step is inflating it in positive and negative direction along the normal of image plane. Here we use an extruding function to control the process of extruding the 2D mesh and to control the displacement of each grid point. 2D Gaussian is a good choice. We tweak the width, height of a standard 2D Gaussian to match the scale of the mesh which corresponded to the original shape to be skeletonized in the image.

In skeleton extraction stage we adopt mesh contraction algorithm developed by Oscar et al. This method works on three dimensional meshes and can extract one dimensional curve skeleton within several contraction iterations. The process of mesh contraction is handled by a discrete Laplace equation. By solving this discrete Laplace equation we get a contracted mesh which is visually thinner than the original mesh. By repeating this process on to the contracted mesh several iterations we would get an extremely contracted 3D mesh witch look like a one dimensional curve skeleton. In Oscar's method a connectivity surgery is applied to the extremely contracted 3D mesh to get a real 1D curve skeleton. The extremely contracted 3D mesh contain a large number of degenerate vertices, so we omit the connectivity surgery step and parallel project the resulting 3D mesh onto image plane directly. Overlapped vertices would be projected to same pixel, and the projected pixels in image space constitute the skeleton representation of the original object or shape.

4 3D Mesh Generation

We use snake algorithm to calculate the contour of an image object. This contour represent the shape of this image object, we considered the problem of skeletonizing a 2D shape as a projection of skeleton extraction of an object in 3D space. We need to reconstruct the 3D geometry of the 2D image object. The reconstruction would be an under-constrained problem because there are many (in fact infinite) 3D shapes satisfy the only constrain of exactly projecting onto the image object. Our 3D mesh generation method does not try to find a best 3D shape by introducing some prior condition or additional constrains. Instead, we just inflate the 2D shape using some weight function. The method is very simple: we first tessellate the 2D shape into a 2D mesh, and then move these mesh vertices off along the normal direction of the image plane.

Our tessellation of an image object is applied by scanning the image object to find out all pixels that are lying on the contour or inside this object. These pixels form uniform rectangle grids if we consider the pixels as vertex and draw edges between one pixel and its neighbor pixels in vertical and horizon direction. This grid can be seemed as quadrilateral mesh and we tessellate it further by adding edges between one pixel and its 45o diagonal neighbor pixel. Then we get a triangle mesh of the image object.

To generate a 3D triangle mesh which is ready to extraction of curve skeleton from it we need a closed surface mesh instead of the mesh we get so far. Our strategy is to clone the mesh and let these two meshes share vertices that are lying on the contour. By doing so, we finally get a flattened closed 3D triangle mesh. The only thing left to do is to inflate it to a real 3D mesh. We move vertices of the flattened 3D mesh in positive and negative direction along the image normal. A weighting function is used to control how far each vertex to move. In this paper, we select a simple strategy that using Gaussian as our weighting function.

5 Curve Skeleton Extraction

Extraction of curve-skeleton is a fundamental problem in computer graphics especially in computer animation. Mesh contraction based curve skeleton extraction contract a 3D mesh to a 1D curve skeleton by applying a implicit Laplacian smoothing with global positional constrains. In general, Laplacian smoothing without constrain would smooth out all detail (as well as features which is important to skeleton extraction) and the resulting 3D mesh would converge into a single point. To preserve skeleton information of the original object, global positional constrains are need to keep features where they are.

Laplacian smoothing is a operation that move every vertices on a mesh by solving discrete Laplacian equation: $\mathbf{LV}'=0$, where $\mathbf{L}$ is a n×n curvature-flow Laplace operator and can be written as follow:

$$\mathbf{L}_{ij}=\begin{cases}\omega_{ij}=cot\alpha_{ij}+cot\beta_{ij} & \text{if } (i,j)\in \mathrm{E}\\ \sum_{(i,k)\in \mathrm{E}}^{k}-\omega_{ik} & \text{if } i=j\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

Where α_{ij} and β_{ij} are the opposite angles with respect to the edge (i, j) [5]. According to [5], $\delta=\mathbf{LV}=[\delta_1^T,\delta_2^T,...,\delta_n^T]^T$ approximate the (inward) curvature-flow normals, and solving equation (1) means removing the normal components and so that contracting the mesh geometry. As descript as above, applying Laplacian smoothing without constrain would make the 3D shape to converge to a single point. So extra global positional constrains are add in [5]. The resulting linear system to be solved is written as below:

$$\begin{bmatrix} \mathbf{W}_L\mathbf{L} \\ \mathbf{W}_H \end{bmatrix} \mathbf{V}' = \begin{bmatrix} 0 \\ \mathbf{W}_H\mathbf{V} \end{bmatrix} \tag{2}$$

Where W_L and W_H are diagonal weighting matrices with respect to Laplacian smoothing term and global positional constrains. The upper half of equation (2) means contract the 3D mesh using Laplacian smoothing and the lower half mean retain the original shape. Equation (2) is over-determined and can be solve in least-squares sense to minimizing the quadratic energy:

$$\|\mathbf{W}_L\mathbf{L}\mathbf{V}'\|^2 + \sum_i \mathbf{W}_{H,i}^2 \|v_i' - v_i\|^2 \tag{3}$$

Solving equation (2) or minimizing (3) we get a shrinked mesh while retain the main shape features compared to the original mesh. By solving equation (2) for the result mesh we get a more shrinked mesh and after several iterations the original mesh can be eventually contracted into a 1D skeleton. We refer readers to the paper [5] for more detail on both math and implementation information.

The resulting 1D skeleton contains a large number of collapsed faces and degenerated faces, so Oscar et al. proposed two post processing step named connectivity surgery and embedding refinement to get a true skeleton structure of the original 3D shape. In this paper, however, we only need a 1D skeleton of the generated 3D shape nevertheless whether it is a really 1D structure or a degenerated geometry shape. Thus we omit the followed steps and stop our skeleton extraction task as we find a visually 1D curve skeleton.

6 Projection

Once we get a visually 1D curve skeleton (although this one is a degenerated 3D shape in fact), we perform a parallel projection from 3D space back to image space along normal of image plane. Because we didn't perform connectivity surgery and embedding refinement to the contracted mesh, so the projection operations are performed to all vertices of the contracted mesh (in which case the number of vertices is the same as original mesh).

Location and orientation of image plane is determined by O_i, $\mathbf{n}$, $\mathbf{x}$ and $\mathbf{y}$. As illustrated in Fig 1., O_i is the origin of image plane, i.e. the left bottom corner of the input image. $\mathbf{n}$ is unit normal vector of image plane. $\mathbf{x}$ and $\mathbf{y}$ are unit vectors along image horizontal and vertical directions. Let P be a vertex on the 1D skeleton, and $p = (P - O_i)$ be the vector from O_i to P, and p' be the projection point of P on image.

Then the projection p' of vertex P can be achieved using equation (4):

$$\mathbf{p'} = \mathbf{p} - \alpha\mathbf{n} \quad \text{where} \quad \alpha = \mathbf{p} \cdot \mathbf{n}; \tag{4}$$

As we get projection point p' of P, the pixel position can be calculated as the dot product between projected point p' and axis of image plane.

$$x = \mathbf{p'} \cdot \mathbf{x}; \qquad y = \mathbf{p'} \cdot \mathbf{y} \tag{5}$$

By labeling all hit pixels we finally get the skeleton of original image object.

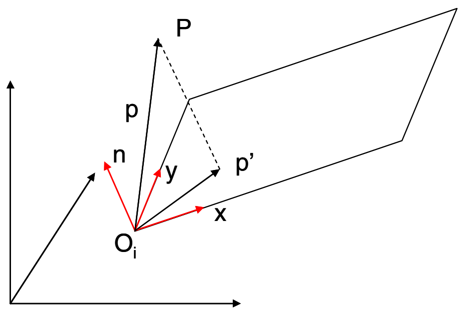


Fig. 1. Project vertex back to image plane: P is a vertex of 1D skeleton of the contracted mesh, it is projected onto image plane and then its image coordinate is calculated with respect to image left-bottom corner O_i

7 Result

Figure 2 shows the original image with a rectangle object in it, the corresponding 3D mesh generated using our method and the resulting skeleton. Figure 3 shows the result of applying our method onto the same rectangle with two little protuberances. We can see that the resulting skeletons are generated correctly even when the boundary is disturbed. Figure 4 and Figure 5 are basically the same as figure 2 and 3, but the objects are thinner.

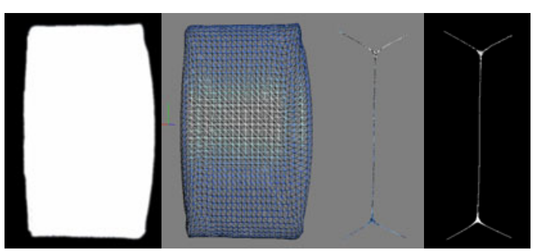


Fig. 2. Skeletonization of a rectangle

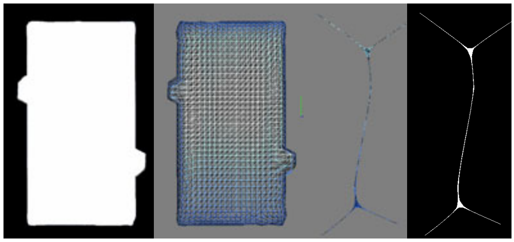


Fig. 3. Skeletonization of a disturbed rectangle

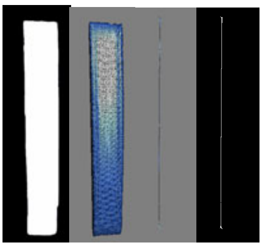


Fig. 4. Skeletonization of thin rectangle

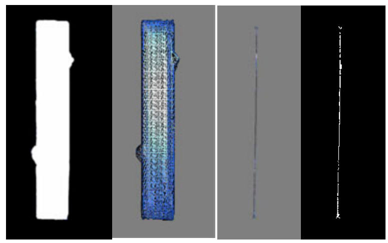


Fig. 5. Skeletonization of disturbed thin rectangle

References

1. Amenta, N., Choi, S., Kolluri, R.K.: The power crust. In: Symposium on Solid Modeling and Applications, pp. 249–266 (2001)
2. Aujay, G., H'Etroy, F., Lazarus, F., Depraz, C.: Harmonic skeleton for realistic character animation. In: Symposium on Computer Animation, pp. 151–160 (2007)
3. Arcelli, C., Sanniti di Baja, G.: A Width Independent Fast Thinning Algorithm. IEEE Trans. PAMI 7, 463–474 (1985)
4. Chen, Y.-S., Yu, Y.-T.: Thinning Approach for Noisy Digital Patterns. Pattern Recogintion 33(5), 809–819 (1996)
5. Au, O.K.-C., Tai, C.-L., Chu, H.-K., Cohen-Or, D., Lee, T.-Y.: Skeleton Extraction by Mesh Contraction. ACM Trans. Graph. 27(3) (2008)
6. Cornea, N.D., Silver, D., Yuan, X., Balasubramanian, R.: Computing hierarchical curve-skeletons of 3D objects. The Visual Computer 21(11), 945–955 (2005)
7. Cornea, N.D., Min, P., Silver, D.: Curve-skeleton properties, applications, and algorithms. IEEE Transactions on Visualization and Computer Graphics 13(3), 530–548 (2007)
8. Dey, T.K., Sun, J.: Defining and computing curve-skeletons with medial geodesic function. In: Symposium on Geometry Processing, pp. 143–152 (2006)
9. Bookstein, F.L.: The Line Skeleton. CGIP: Comp. Graphics and Image Proc. 11, 123–137 (1979)
10. Leymarie, F., Levine, M.D.: Simulating the Grass fire Transform Using an Active Contour Model. IEEE Trans. PAMI 14(1), 56–75 (1992)
11. Geiger, D., Liu, T.-L., Kohn, R.V.: Reprensentation and self-similarity of shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(1), 86–99 (2003)

12. Giblin, P.J., Kimia, B.B.: On the intrinsic reconstruction of shape from its symmetries. IEEE Trans. PAMI 25(7), 895–911 (2003)
13. Blum, H.: A Transformation for Extracting New Descriptors of Shape. In: Models of the Perception of Speech and Isual Form. MIT Press, Cambridge (1967)
14. Hassouna, M.S., Farag, A.A.: Robust centerline extraction framework using level sets. In: Proceedings on Computer Vision and Pattern Recognition, vol. 1, pp. 458–465 (2005)
15. Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L.: Topology matching for fully automatic similarity estimation of 3D shapes. ACM Trans. Graph., 203–212 (2001)
16. Brandt, J.W., Algazi, V.R.: Continuous Skeleton Computation by Voronoi Diagram. CVGIP: Image Understanding 55, 329–338 (1994)
17. Ma, C.-M., Wan, S.-Y., Lee, J.-D.: Three dimensional topology preserving reduction on the 4-subfields. IEEE Transactions on Pattern Analysis and Machine Intelligence 4(12), 1594–1605 (2002)
18. Ma, W.-C., Wu, F.-C., Ouhyoung, M.: Skeleton extraction of 3D objects with radial basis functions. In: Proceedings of the Shape Modeling International, pp. 207–215 (2003)
19. Ogniewicz, R., Ilg, M.: Voronoi skeletons: Theory and applications. In: Proceedings on Computer Vision and Pattern Recognition, pp. 63–69 (1992)
20. Pascucci, V., Scorzelli, G., Bremer, P.-T., Mascarenhas, A.: Robust on-line computation of Reeb graphs: simplicity and speed. ACM Trans. Graph. 26(3), article 58 (2007)
21. Palágyi, K., Kuba, A.: A parallel 3D 12-subiteration thinning algorithm. Graph. Models Image Process. 61(4), 199–221 (1999)
22. Kimmel, R., et al.: Skeletonization via Distance Maps and Level Sets. CVIU: Comp. Vision and Image Understanding 62(3), 382–391 (1995)
23. Ogniewicz, R.L.: 1993. Discrete Voronoi Skeletons. Hartung-Gore (1993)
24. Shi, F.Y., Wong, W.-T.: Fully Parallel Thinning with Tolerance to Boundary Noise. Pattern Recogintion 27(12), 1677–1695 (1994)
25. Tang, Y.Y., You, X.: Skeletonization of ribbon like shapes based on a new wavelet function. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1118–1133 (2003)
26. Liu, T.L., Geiger, D., Kohn, R.V.: 1999. Representation and Self Similarity of Shapes. In: Proc. ICCV 1999, pp. 1129–1135 (1999)
27. Wang, Y.-S., Lee, T.-Y.: Curve skeleton extraction using iterative least squares optimization. IEEE Transactions on Visualization and Computer Graphics (2008)
28. Zhou, Y., Toga, A.W.: Efficient skeletonization of volumetric objects. IEEE Transactions on Visualization and Computer Graphics 5(3), 196–209 (1999)

Appearance Similarity Index for Medicinal Ampoule Labels

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

[1]Shibaura Institute of Technology,
3-7-5 Toyosu, Koto Ward, Tokyo 135-8548, Japan
{masaomi,ohkura}@shibaura-it.ac.jp
[2]Graduate School of Technology, Shibaura Institute of Technology,
3-7-5 Toyosu, Koto Ward, Tokyo 135-8548, Japan
{m110013,m709102}@shibaura-it.ac.jp
[3]International University of Health and Welfare,
2800-1 Kitakanemaru, Ohdawara City, Tochigi 324-8501, Japan
ftsuchiya@iuhw.ac.jp

Abstract. Since there are many ampoule injection medicines, it is important to make their labels easily distinguishable because confusing labels may lead to fatal accidents caused by administering the wrong medicine by mistake. In this paper, we utilize Fourier series expansion and wavelet transformation to extract the characteristics in labels and propose an index to measure similarity that we feel toward ampoule labels to prevent confusion in label designs. We also discuss a way of parameterizing colors.

Keywords: Medicinal safety, Ampoule labels, Fourier analysis, Wavelet analysis.

1 Introduction

Since there are many ampoule injection medicines, it is important to make their labels easily distinguishable because confusing labels may lead to fatal accidents caused by administering the wrong medicine by mistake. However, labels tend to be similar because they are small and there is little freedom in their design elements. It is necessary to define an appearance similarity index and to warn authorities against approving labels that may cause confusion.

The confusion over labels is similar to the problem of medicines having similar names, since this can also result in accidents caused by administering the wrong medicine by mistake. Tsuchiya et al. [1] proposed indices that measure the similarity of medicine names, such as the cosine value, which measures the angle between vectors whose elements denote the frequency of each letter. He also developed a similar name search system based on the indices. The system is actually used to prevent approval of medicine that has a similar name to those of existing medicines. However, there have been no studies on evaluating the visual similarities of packages, especially ampoules, of injection medicines. Nabeta and Imai et al. [2] proposed a

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 588–597, 2011.
© Springer-Verlag Berlin Heidelberg 2011

similarity index that takes account of the similarity of letters appearing in medicine names.

Although the color and material of the container part can contribute to their similarity, in this paper, we focus on the similarity of ampoule labels because there is less variety in the color and material of container parts than labels and because how to measure similarity that we feel toward ampoule labels has not been clarified.

In order to design a similarity index for ampoule labels, we discuss how to extract data that contain information of their characteristics. In this paper, we introduce two ways of extracting such data: Fourier series expansion and wavelet transformation.

Fourier series expansion is useful for finding characteristics of periodic functions. Since the cross-section of labels forms a circle, the distribution of colors on labels can be regarded as an angular distribution function, which is periodic. Namely, if you deal with the components obtained by Fourier series expansion, it is automatically guaranteed that the original data is periodical. Therefore, we do not directly treat the original color distribution but the Fourier coefficients as the fundamental characteristic values to represent the angular color distribution of labels.

We also introduce analysis that utilizes (discrete) wavelet transformation[3] instead of Fourier series decomposition because we can feel similarity of not only the whole part of each label but also the part that is within the range of view. Although Fourier series decomposition provides us with the characteristics data of the whole label, wavelet transformation shows components of the localized 'wave' (wavelet) in the target function. We expect that the similarity of the components corresponds to similarity in the range of view.

We also discuss a way of parameterizing colors. Red, Green and Blue (RGB) are the most common color space. In this paper, we show that labels that have a wide white area in the background tend to have higher similarity values to any other labels. (Note that most labels have a white background.) We also adopt other color spaces, L*a*b* in order to decompose the lightness elements from the color space.

Based on these findings, we propose an appearance similarity index of labels. We regard the subset of the coefficient series obtained from Fourier analysis or wavelet analysis as a vector, and calculate cosine values that measure the angles between each pair of such vectors.

We evaluate our index by applying it to realistic ampoule labels. We calculate the values of our index and compare them with the answers of questionnaires asking the extent of similarity.

2 Methods

2.1 The Method Based on Fourier Series Expansion

In general, we utilize Fourier series expansion,

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \sin n\theta + a_{-n} \cos n\theta) \quad (1)$$

in order to find the characteristics of periodic functions. In this study, we can regard the distribution of colors on ampoule labels as an angular distribution function along the cross-section of ampoules. The function is periodic since the cross-section of ampoules is a circle. Based on this consideration, we treat not the original color distribution but the Fourier coefficients as the fundamental data.

Let us parameterize each pixel of a label by a set of values (τ, θ), where θ denotes the angle measured along the edge of a cross-section of an ampoule and denotes the coordinate that is along the axial direction of the ampoule. Let R_θ^τ, G_θ^τ, B_θ^τ denote the color component values of red, green, blue respectively at the pixel (τ, θ). If we apply Fourier series decomposition to them, we find that

$$\begin{pmatrix} R_\theta^\tau \\ G_\theta^\tau \\ B_\theta^\tau \end{pmatrix} = \begin{pmatrix} r_0^\tau \\ g_0^\tau \\ b_0^\tau \end{pmatrix} + \sum_{n=1}^{\infty} \left\{ \begin{pmatrix} r_n^\tau \\ g_n^\tau \\ b_n^\tau \end{pmatrix} \cos n\theta + \begin{pmatrix} r_{-n}^\tau \\ g_{-n}^\tau \\ b_{-n}^\tau \end{pmatrix} \sin n\theta \right\}. \tag{2}$$

We deal with the first term and the other terms separately because the first term is a set of the mean values of each color and the other terms contain information of the extent of variation from the mean values of each color.

From now on, we assume that people feel similarity of labels based on the color and the display design (w/o color) printed on labels.

As for the mean value components, the corresponding spectral density at τ for each color is given by:

$$\frac{(r_0^\tau)^2}{2}, \frac{(g_0^\tau)^2}{2}, \frac{(b_0^\tau)^2}{2} \tag{3}$$

Since each of these represents the typical magnitude of each color at , we regard

$$v_\tau = \left(\frac{(r_0^\tau)^2}{2}, \frac{(g_0^\tau)^2}{2}, \frac{(b_0^\tau)^2}{2} \right) \tag{4}$$

as a vector whose direction indicates the tendency of color at τ . For the pair of drugs, we calculate the average similarity of colors as

$$sim^{color} = \frac{1}{m} \sum_{\tau} \frac{v_\tau \cdot v'_\tau}{\|v_\tau\| \|v'_\tau\|} \tag{5}$$

where m is the length (the number of pixels) of a label in the direction of τ .

Remember that $r_{\pm n}^\tau$, $g_{\pm n}^\tau$, $b_{\pm n}^\tau$ represent the magnitude of variation. We expect that these values contain information on display designs other than colors, such as lines, marks, letters and so forth. We assume that labels that have analogous patterns make us feel similarity even if their color is different. In order to extract the pattern information, we convert the color components R_θ^τ, G_θ^τ, B_θ^τ to grayscale, based on the approximate expression:

$$\mathrm{Gray}_\theta^\tau = \frac{2R_\theta^\tau + 4G_\theta^\tau + B_\theta^\tau}{7} \tag{6}$$

$$= \frac{2r_0^\tau + 4g_0^\tau + b_0^\tau}{7} + \sum_{n=1}^{\infty} \left(\frac{2r_n^\tau + 4g_n^\tau + b_n^\tau}{7} \sin n\theta + \frac{2r_{-n}^\tau + 4g_{-n}^\tau + b_{-n}^\tau}{7} \cos n\theta \right)$$

The components of their spectral density are given as:

$$q_n^T = \left(\frac{2r_n^T + 4g_n^T + b_n^T}{7}\right)^2 + \left(\frac{2r_{-n}^T + 4g_{-n}^T + b_{-n}^T}{7}\right)^2. \tag{7}$$

We regard a series of these as a vector and define the average similarity of display patterns as follows:

$$sim^{design} = \operatorname*{avg}_{T} \frac{\sum_n q_n^T q_n'^T}{\sqrt{\sum_n q_n^{T2}}\sqrt{\sum_n q_n'^{T2}}} \tag{8}$$

Now we have two similarity indices and find the way to unify these. We investigated the distributions of both similarity indices calculated for the pairs of 12 ampoule labels available in the market. The results are shown in Fig.1.

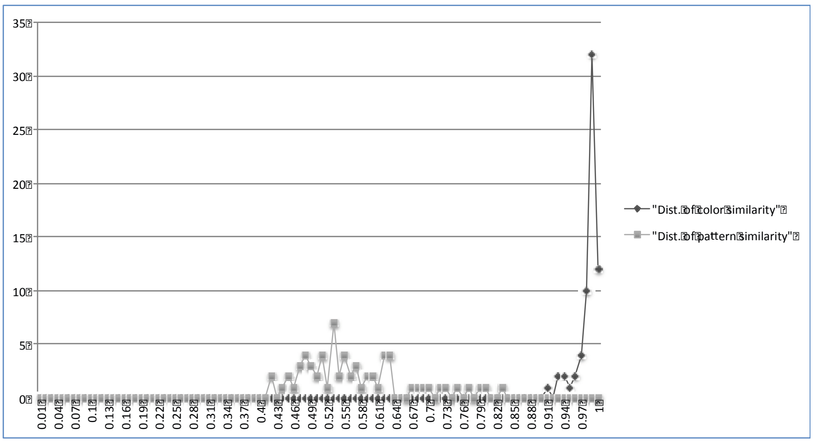


Fig. 1. Distributions of similarity indices

This figure tells us that the color similarity index shows unimodal distribution although the pattern similarity index shows multimodal distribution. This difference cannot be absorbed even by normalization. We focus on their rank to preserve the order of similarity. Let $rank_c$ and $rank_p$ denote the functions that return the rank counted in the ascendant order. The suffix c and p denotes the rank of color similarity index and pattern index, respectively. If N is the number of target ampoule labels, the maximum value of $rank$ is the number of the combinations of labels, namely $(N-1)/2$. Based on these considerations, we defined the normalized unified similarity as:

$$sim^f = 1 - \frac{rank_c(sim^{color}) + rank_p(sim^{design}) - 2}{N(N-1)} \tag{9}$$

2.2 The Method Based on Wavelet Transformation

We also introduce analysis that utilizes (discrete) wavelet transformation instead of Fourier series decomposition. This is under the assumption that we feel similarity of not only the whole part of each label but also the part that is within the range of view. Although Fourier series decomposition provides us with the characteristics data of the whole label, wavelet transformation shows the characteristic data of the parts covered by the localized 'wave' (wavelet) in the target function. We expect that the similarity of the parts corresponds to similarity in the range of view.

Let $h^{\tau}(x)$ be a target color component function at τ , where $x = \frac{\theta}{2\pi}$. Remember that the colors printed on the typical labels do not change gradationally. Because of this, we use Haar's mother wavelet defined as

$$\psi(x) = \begin{cases} 1 & 0 < x \leq \frac{1}{2}, \\ -1 & \frac{1}{2} < x \leq 1, \\ 0 & otherwise. \end{cases} \tag{10}$$

In general, $h^{\tau}(x)$ can be decomposed as

$$h^{\tau}(x) = h^{\tau}_{J}(x) + \sum_{l=J}^{\infty} \sum_{k} \gamma^{\tau}_{l,k} \psi(2^{l} x - k), \tag{11}$$

where l is a scaling parameter and k denotes the displacement of a translation of Haar's wavelet. The function $h^{\tau}_{J}(x)$ is the step function approximation of $h^{\tau}(x)$, each of whose intervals has the width 2^{-J}.

We assume that the display printed on each 1/4 part of a label can be recognized at a time, and each half of 1/4 part is overlapped. Taking account of these assumptions and the size of letters on the labels, we approximate $h^{\tau}(x)$ as

$$h^{\tau}(x) \approx h^{\tau}_{3}(x) + \sum_{l=3}^{5} \sum_{k} \gamma^{\tau}_{l,k} \psi(2^{l} x - k). \tag{12}$$

Let $\mu^{\tau}_{j} = h^{\tau}_{3}(j/8)$, which is equal to the average of $h^{\tau}(x)$ in the range of $j/8 \leq x < (j+1)/8$. We measure the similarity based on μ^{τ}_{j} and $\gamma^{\tau}_{l,k}$ not $h^{\tau}(x)$ itself. We define the vector $\xi^{\tau}_{j} = \left(\mu^{\tau}_{j}, \gamma^{\tau}_{3,j}, \gamma^{\tau}_{4,2j}, \gamma^{\tau}_{4,2j+1}, \gamma^{\tau}_{5,4j}, \gamma^{\tau}_{5,4j+1}, \gamma^{\tau}_{5,4j+2}, \gamma^{\tau}_{5,4j+3} \right)$ for $0 \leq j \leq 7$.

Although we used the RGB color model in Section 2.1, we use the L*a*b* color model. The value of L* represents the lightness of the color, a* represents the position between red and green and b* represents the position between yellow and blue. In fact, most labels have a white background, which is one reason why the labels resemble each other. The reason for using this model is that it is easy to separate the degree of freedom related to white, namely, lightness L*. In order to neglect this degree of freedom, we set $\mu^{\tau}_{j} = 0$ for L*.

Finally, we join the vectors by direct sum:

$$\rho_j^\tau = (\xi^L)_j^\tau \oplus (\xi^a)_j^\tau \oplus (\xi^b)_j^\tau,$$

and define the average similarity of the parts of labels as

$$sim^W_{j,j'} = \operatorname{avg}_\tau \frac{\rho_j^\tau \cdot \rho'^\tau_{j'}}{\|\rho_j^\tau\| \|\rho'^\tau_{j'}\|},$$

and the similarity of the whole labels as

$$sim^W = \max_k \frac{1}{8} \sum_j sim^W_{j,j+k}.$$

3 Experiments

We estimated our methods by applying them to the ampoule labels provided in the market. The targets were the 12 labels shown in Fig. 2. Since each of the label images has a different size, we adjusted the size to 157-pixel height and 256-pixel width by means of linear interpolation method.

3.1 sim^f

Fig.2 shows the pairs of labels and their values of similarity index sim^f in descending order. The pair, Label6 and Label7, and the pair, Label8 and Label9, are respectively very similar designs that have the same colors. The pair, Label5 and Label6, and the pair, Label5 and Label8, have similar design but different colors, and their values of sim^f are less than those of the above pairs. These facts are compatible with our intuition.

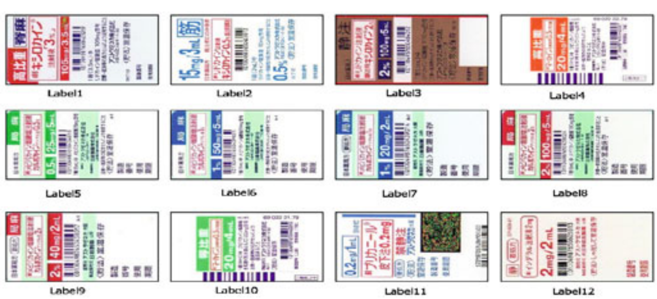


Fig. 2. The sample labels for the experiment

Table 1. sim^f for the pairs of the labels in Fig.2

Order	Labels		Color		Design w/o color		sim^f
	X	Y	sim^{color}	rank	$sim^{pattern}$	rank	
1	6	7	0.9990	1	0.7165	7	0.9546
2	8	9	0.9989	2	0.7303	6	0.9546
3	9	12	0.9979	3	0.6328	13	0.8939
4	5	6	0.9945	14	0.7678	4	0.8788
5	5	8	0.9936	18	0.8004	2	0.8636
6	9	11	0.9934	20	0.6345	12	0.8409
7	4	9	0.9970	6	0.5717	27	0.7576
8	8	11	0.9939	17	0.6208	17	0.75
9	5	7	0.9919	26	0.6984	8	0.7424
10	2	6	0.9977	4	0.5523	31	0.7424
⋮							
62	3	11	0.9457	61	0.5287	41	0.2424
63	10	12	0.9874	41	0.4372	64	0.2197
64	3	12	0.9679	56	0.5038	44	0.2121
65	1	10	0.9835	49	0.4422	63	0.1667
66	3	10	0.9363	63	0.4492	62	0.0682

So are the pairs that have small sim^f values. This is also compatible with our intuition, since their design and color is different.

The pair, Label9 and Label12, also has a large sim^f value, though they can be intuitively recognized to be different. In detail, we can see that their base colors are red and that the position and size of the bar code are similar. In fact, Label12 has a relatively large sim^{design} value with Label5, Label6, Label7, Label8 and Label9. The void (white) parts at their upper right are wider than the others and therefore have large RGB values in large areas. The pattern of the void parts makes their $sim^{pattern}$ value large. In spite of these facts, they seem to be different. This suggests that the contribution of the void parts is larger than the contribution of different parts between labels.

3.2 sim^w

The pairs that have high sim^w value are the combinations of Label5, Label6, Label7, Label8 and Label9. It is easy to see that the results reflect the difference of design such as the width of colored stripes with standard units (e.g. 1% 20 mg/2 ml) rather than their colors. We consider that this comes from the characteristics of the multi-resolution analysis (MRA) based on Haar's wavelets, since MRA measures the scale of the stripe width.

The exception is the pair, Label1 and Label2. It is notable that the letters in the right half of Label1 are printed only in black and the right half of Label2 has a similar layout of letters apart from their color (blue). We should remember that black and white have only L* values (namely a* and b* are zero) and that blue color has a non-zero L* value. Since we defined the similarity based on an inner product of the vectors ρ_j^r, and since $\xi_j^{ar} = 0$ and $\xi_j^{br} = 0$ for the right half of Label1, if the vectors ξ_j^{Lr} for the two labels are almost parallel, sim^w has a large value. This condition means that the labels have a similar design.

In this experiment, the average of $sim_{j,j+k}^W$ for the pair, Label5 and Label10, was maximized to be 0.1951 when the parameter k was 1. Although both of these have similar green parts, the positions where green parts exist in the labels have minor deviation. The result k=1 indicates that shifting Label10 to the left as much as 256/8=32 pixels makes the labels look most similar.

Table 2. sim^w for the pairs of the labels in Fig.2

Order	Labels		sim^W
	X	Y	
1	5	6	0.4250
2	6	8	0.4198
3	7	9	0.4070
4	5	8	0.3715
5	6	9	0.3192
6	8	9	0.2803
7	1	2	0.2471
8	5	9	0.2283
9	7	8	0.2235
10	6	7	0.2076
⋮			
62	2	5	0.02837
63	1	10	0.01114
64	3	7	0.006361
65	2	8	0.004041
66	4	7	0.000142

3.3 Comparison of sim^f and sim^w with the Results of a Questionnaire Survey

We conducted a questionnaire survey to compare our similarity indices with the subjective estimation of similarities answered by respondents. We presented two pairs of labels to respondents and asked them to answer which pair was more similar than the other.

The respondents were 10 university students.

For both sim^f and sim^W, the pair that has a higher similarity value is more selected by respondents than the other in about 72% of cases (8 cases out of 11 cases). Especially for the cases that a pair has $sim^W \geq 0.10$ the other has $sim^W < 0.10$, more than 90% of respondents tended to answer that the former pair is similar. This suggests that our similarity indices tend to coincide with the similarity felt by humans.

As we explained, the void part in the background of labels tends to enlarge the value of sim^f . The result suggests that this holds for Case 2, Case 3 and Case 4.

As for sim^w , Label3 tends to have large values with other labels as is seen in Case 6 and Case 8. This is because sim^w ignores the brightness/darkness of the background and tends to focus on the design rather than colors. This can cause a high value of sim^w even though the background color is different.

These facts suggest that the key is how to deal with bulk (background) colors of labels and reflect them in similarity indices.

Table 3. Comparison of sim^f and sim^w with the results of the questionnaire

Case	Combination A				Combination B			
	labels	sim^f	sim^w	#res	labels	sim^f	sim^w	#res
1	*2,6*	**0.75**	**0.063**	**6**	*7,11*	0.58	0.041	4
2	*2,6*	**0.75**	0.063	1	*5,10*	0.55	**0.20**	**9**
3	*1,8*	0.58	0.035	**7**	*9,12*	**0.89**	**0.11**	3
4	*9,11*	**0.77**	0.039	2	*4,10*	0.59	**0.16**	**8**
5	*4,7*	0.40	0.00014	4	*5,11*	**0.64**	**0.039**	**6**
6	*3,8*	0.45	**0.081**	1	*1,12*	**0.51**	0.056	**9**
7	*2,7*	0.62	0.068	2	*7,8*	**0.64**	**0.22**	**8**
8	*4,9*	**0.77**	0.082	**9**	*3,6*	0.36	**0.11**	1
9	*4,12*	0.56	0.096	1	*6,8*	**0.72**	**0.42**	**9**
10	*9,12*	**0.89**	**0.11**	**10**	*4,5*	0.41	0.072	0
11	*2,10*	0.41	0.034	1	*5,6*	**0.88**	**0.43**	**9**

4 Conclusion

In order to design a similarity index for ampoule labels, we discussed how to extract data that contain the information of their characteristics. In this paper, we introduced methods based on Fourier series expansion and wavelet transformation.

As for Fourier series expansion, we separately dealt with the constant component and other components. This is because the constant term is a set of the mean values of each color and the other terms contain information of the extent of variation from the mean values of each color. We defined a similarity index for each of these components and combined them by means of their ranks in order to absorb the difference of their distribution.

We also introduced analysis that utilizes wavelet transformation based on Haar's wavelet in order to estimate similarity of not only the whole part of each label but also its part that is within the range of view. We defined a similarity index based on the cosine given by the inner product of the vectors whose elements are wavelet coefficients.

We also discussed the color parameterization. For Fourier analysis, we used the RGB color system, which is the most commonly used. In this paper, for the wavelet analysis, we also adopted L*a*b* in order to decompose the lightness elements from the color space.

We evaluated our methods by applying them to the ampoule labels provided in the market and compared the obtained similarity index values with the subjective estimation of similarities responded by examinees.

For both similarity indices, the pair that has a higher similarity value is more selected by respondents than the other in about 72% of cases (8 cases out of 11 cases).

The void part in the background of labels tends to enlarge the value of the similarity index based on Fourier series decomposition. The similarity index based on our wavelet analysis ignores the brightness/darkness of the background and tends to focus on the design rather than colors. These facts suggest that the key is how to deal with bulk (background) colors of labels to reflect them in similarity indices.

In the next step, we will propose a method of dealing with the background color of labels. It is not straightforward, since white background color is too common to be the characteristic of labels though the other background color is sensitively recognized to be a characteristic.

Although we ignored the color, size and material of the container part of ampoules, they can have a large impact on similarity. In the future, we will combine their contributions with our method.

References

1. Tsuchiya, F., Kawamura, N., Oh, C., Hara, A.: Standardization and similarity deliberation of drug-names. Japan Journal of Medical Informatics 21(1), 59–67 (2001)
2. Nabeta, K., Imai, T., Kimura, M., Ohkura, M., Tsuchiya, F.: The similarity index of medicine names to prevent confusion. In: Proceedings of Pan-Pacific Conference on Ergonomics 2010, Taiwan (2010)
3. Daubechies, I.: Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, Philadelphia (1992)

A Novel Parallel Clustering Algorithm Based on Artificial Immune Network Using nVidia CUDA Framework

Ruiyi Luo and Qian Yin*

Image Processing and Pattern Recognition Laboratory
Beijing Normal University, Beijing 100875, China
yinqian@bnu.edu.cn

Abstract. In this paper, a novel parallel data clustering algorithm based on artificial immune network aiNet is proposed to improve its efficiency. In consideration of the restrictions of GPU, we carefully designed the data structure, algorithm flow and memory allocation strategy of the parallel algorithm and realized it using NVIDIA's CUDA framework. During the implementation, in order to fully explore its implicit parallelism, we allocated threads on GPU that represent the network cells of aiNet, and modified this algorithm to let those thread operations parallel during the clustering process. We calculated the affinity parallel, combined the random numbers with the local search algorithm to select the first n cell parallel, and did the network suppression parallel. Experimental results show that certain speedup can be obtained by using the proposed method.

Keywords: artificial immune network, aiNet, clustering, parallel, GPU.

1 Introduction

Artificial immune systems have drawn much attention in recent years. Many interesting algorithmic solutions were proposed in some fields, such as clustering or classification [1][2] in data mining, intrusion detection of network [3], optimization [4] and etc. Clustering is an unsupervised classification, which is very useful for data analysis [5][6]. The most popular algorithm for data clustering is K-Means [7]. Though the speed of K-Means is usually very fast, a main drawback of it is that it needs to know the cluster number previously and the result is sensitive to the initial center. In contrast, the artificial immune network clustering is a dynamic clustering algorithm, which means that it can get the clustering number in the analysis process. However, the efficiency of it is lower, and this is because that there are too many calculation steps during the process. The objective of our study is to parallel the aiNet (one of artificial immune network clustering) to improve its efficiency. As we known, the biological immune system is obviously a parallel system to protect the vertebrate's organisms, that's to say, the algorithm that simulates it has inherent parallelization and can be explored. Graphic Processing Unit (GPU) is a highly parallel multithread and multi-core

* Corresponding author.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 598–607, 2011.
© Springer-Verlag Berlin Heidelberg 2011

processor originally designed for accelerating 3D rendering in graphic applications, where need paralleling data processing for speed. It has emerging as one of the most powerful and common parallel processing devices, we also choose it as our parallel tool.

The paper is organized as follows. In section 2, artificial immune network algorithm used in data clustering is introduced, and time complexity analysis is also given. In section 3, we describe the relevant architectural features of GPU and Compute Unified Device Architecture (CUDA) framework. The Implementation of the parallel algorithms using CUDA framework on GPU is proposed. Section 4 shows the empirical results obtained. Section 5 concludes this paper and gives some guidelines for the future work.

2 The aiNet Algorithm

Since the immune network theory had been proposed by Jerne in 1974 [8] and the colonel selection and affinity maturation algorithms proposed by Burent in 1987 [9], there are several useful principles which can be simulated in engineering research were emerged in human immune system. The two main principles of which are as follows. One is the colonel selection principle [10], which explains how the immune system recognizes and reacts to antigen(Ag). The second is the immune network theory[8], which hypothesis a new view point of lymphocyte activities, memory cells, internal images, antigen, antibody(Ab), size control, affinity, etc. In this paper, we choose the aiNet [1] algorithm to parallel for data clustering, which is based on the theory of artificial immune network .

The aiNet algorithm was proposed by De Castro in 2000 and has been described detailed in reference [1]. However, in order to find out which part of this algorithm could be paralleled to improve its efficiency, we find out the most time-consuming processes of the aiNet algorithm and try to execute them parallel. Firstly, In each iteration process, all network cells do the same thing to the antigen presented when calculate affinity, so the time consumed in those steps should be the time one network cell consumed multiply by the size of the network. Secondly, in the selection process, the time of selecting the first N cells takes no less than $o(N * \log(N))$. Thirdly, in network suppression process or clone suppression process, similarity between memory cells is measured, and this step takes no less than$o(N_Size^2)$ (N_Size is the size of the network), which are very time-consuming especially when the N_Size is rather large.

3 The Parallel Algorithm

3.1 Data Structure and the Flowchart

Parallel technology has been widely used in various algorithms to improve the efficiency [11][12][13]. GPU is a highly parallel multithread and multiprocessors which is excellent at floating point operation and parallel computing. It is very appropriate to solve problems that can be expressed as data-parallel, as the same instructions are executed for each data element. There are some other algorithms have been paralleled by researchers, such as the clone selection [14], resource limited artificial immune system [15], and Genetic Algorithm have been carried out on GPU [16]. However, those algorithms are mainly based on the colonel selection principle [10], and our work in this paper is to parallel the immune network system for data clustering based on immune network theory [8].

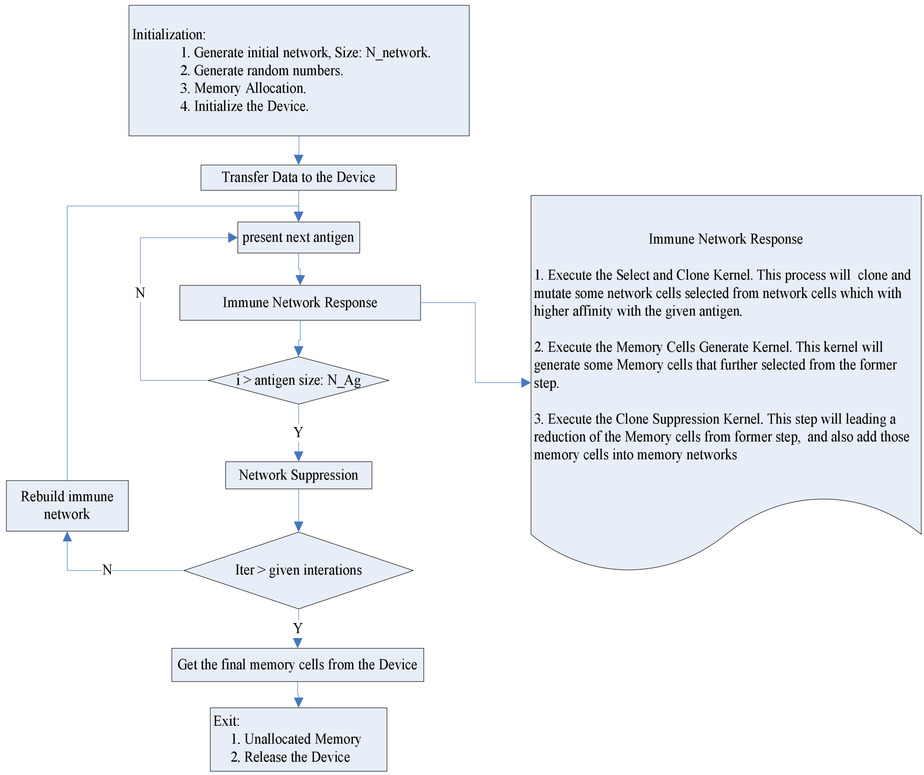


Fig. 1. Flow chart of the proposed algorithm

To implement our parallel algorithm on GPU, we first design the data structure, which takes shape logically in form of a two-dimensional grid plane and is allocated linear storage in shared memory of GPU. A grid of it represents a cell and has four fields: Ab, Affinity, Is_Deleted, Is_HillClimbing. The Ab represents the vector of length n, which is the measurement of the antigen, and it represents the data information that needs for clustering in the real world. The Affinity is a float data which store the affinity value of the cell with the other cells. Is_Deleted indicates whether this cell has been deleted from the network. Is_HillClimbing represents whether we will execute hill climbing process with the cell. More details will be introduced in the following sections. Firstly, we distribute network cells on shared memory and allocate memory on GPU. Then, in each iteration process, for the antigen presents serially to the network, we start the immune response where threads are running parallel. Finally, the network will be rebuilt and the iteration process will be continued. The whole algorithm we proposed briefly works as Fig. 1.

We use the random numbers for the data initialization which are generated by a Kernel introduced in 3.4, and the immune network response process which contains two other kernels will be presented in 3.2 in detail.

3.2 Immune Network Response

The general flow has been introduced in the last section. In this section, more details of immune network response are introduced, which are shown in Fig.2.

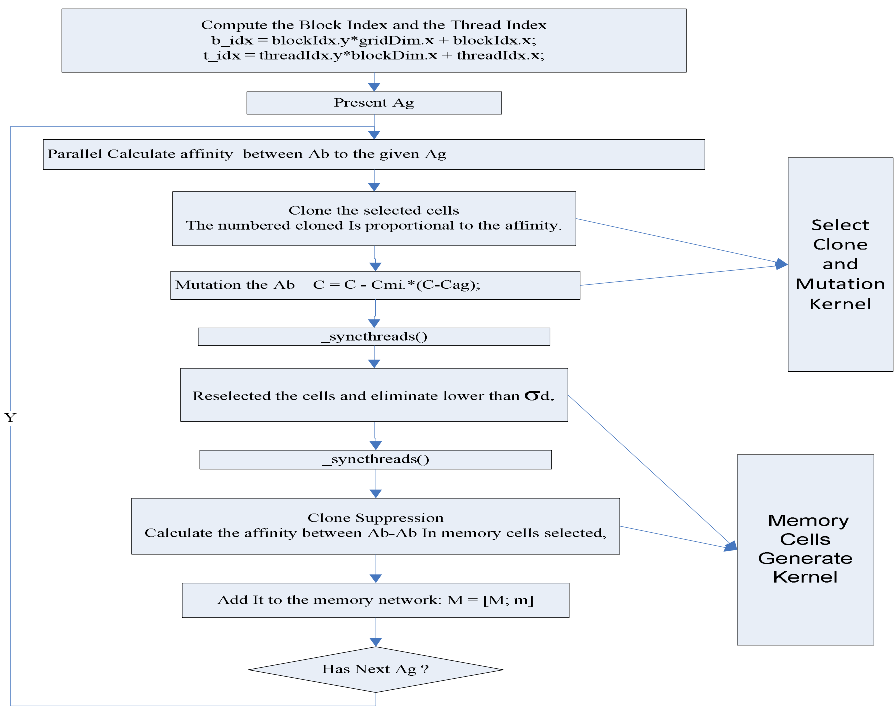


Fig. 2. Immune Response Flowchart

Corresponding to the Fig.2, for the antigen present serially, algorithm we proposed works as follows in detail.

- carry out the step 1 and step 2.
  - Step1: Execute the Select and Clone Kernel: calculate its affinity to all the network cells and select the n highest ones, then clone and mutate them. The higher the affinity, the larger the clone size is.
  - Step2: Carry out Memory Cells Generate Kernel, which recalculates the affinity of the selected cells by step1 with the antigen presented and eliminate those inferior to the threshold δ_d. And reselect the $\xi\%$ most improved cells sorted by affinity. This is also being executed on GPU parallel. Then, execute the clone suppression during which eliminate

those cells whose affinity (Ab-Ab) is inferior to threshold. The memory cells generated are put in memory network.

- Combine the memory cells with the network cells, and execute the Network Suppression Kernel: calculate the whole network inter-cell affinities and eliminate those cells whose affinity with each other is inferior to a given thresholdδ_s; Replace the whose individuals by novel randomly generated ones;

Among the implementation, one of the most challenges is the memory allocation. Firstly, memories are allocated with the size of the network cells on shared memory of GPU to initialize our algorithm; also we allocate some more memories for amplification as after the outer iteration the size of the network will increase. However, the max size is the sum of the initial network size added with the size of Memory Network Cell. For the Clone Network, the max size is equal to the Clone Number Multiplier (a pre-set number) multiplied by the number of threads selected to execute the hill climbing algorithm.

There is also a Network Suppression Kernel like the Clone Suppression Kernel, which is executed after the antigens all have been represented once. For each antigen, the whole network response to it parallel, each thread calculates the affinity, then, judged by itself whether started the local search, and the Is_HillClimbing is set before current iteration started. Euclidean distance is used to calculate the affinity between Ag and Ab or between Ab and Ab.

3.3 Generation of Random Number

The CUDA haven't provided any Random number generator. Here we use the example named MersenneTwister provided by the SDK of CUDA, and the example is based on the algorithm proposed by Makoto Matsumoto and Takuji Nishimura. We need random numbers in the following process: Firstly, we initialize the network cells with the random numbers, which is equal with: N_Network_Size (the size of the network). Secondly, when we decide which thread will continue with the hill climbing local search, some random numbers are also needed, which are transformed into the subscript of the network cells and thread.

$$\xi * \sum_{i}^{N\_Selected} affinity_i * CloneMultiple. \quad (1)$$

What's more, we need to carry out the local search twice during the total process. In the first search, the numbers we need is N_Selected (a pre-set number). In the second search, the numbers we need is calculated as formula (1). We select $\xi\%$ higher affinity cells and clone them. The number of clone size is equal with the value of affinity multiple by CloneMultiple (a pre-set value). Thirdly, when we rebuild the network, we need random numbers to replace the weakest ones in every interaction, and the number of random numbers we needed are N_Network_Size.

The total max number of random we need is as formula (2).

As the $affinity_i$ of each selected cells is not same, it cause that the N_{Random} is not a certain data. However, we use $affinity_{max}$ which is the max affinity value among those cells instead of different $affinity_i$ to get a certain N_{Random}. This generated process is executed once at the start of iteration.

$$N_{Random} = N_{Network_{Size}} + N_{Selected} + \xi\% * \sum_{i}^{N\_Selected} affinity_i * CloneMultiple + N\_Network\_Size. \quad (2)$$

3.4 Parallel Local Search

In the proposed algorithm, we use the hill climbing to search the first n affinity cells. To make the program continuous and avoid switch threads, we use random numbers to decide which thread will execute this process. If the thread starts, it firstly selects the one with the highest affinity from its four sides in the data structure: left, right, up, down. Then if the affinity of the center cell where the thread start is larger than the other four, the thread will terminate. Otherwise, the highest one will be the center and continue the same operation. In this process, in order to avoid that all the threads get one highest cell but the n (N_Selected) first, the number of iterations (ITER_HillClimbing) is set a certain value based on the ratio of n and total network cells number. The left part of Fig.3 has shown a thread that carries out local search of the hill climbing. If the ITER_HillClimbing we set was small, it would only search for a smaller scope, such as in our grid data structure. If we set ITER_HillClimbing to three, most eleven grids we could search are around the center blue cell. As the right part of Fig.3 shown, the empty grids represent there are no cells generated or have been deleted, we will take c as the below cells of A. If we first get a center with B, we will also search its four sides, and then we get b in the first iteration, and then start the next iteration from the center of b.

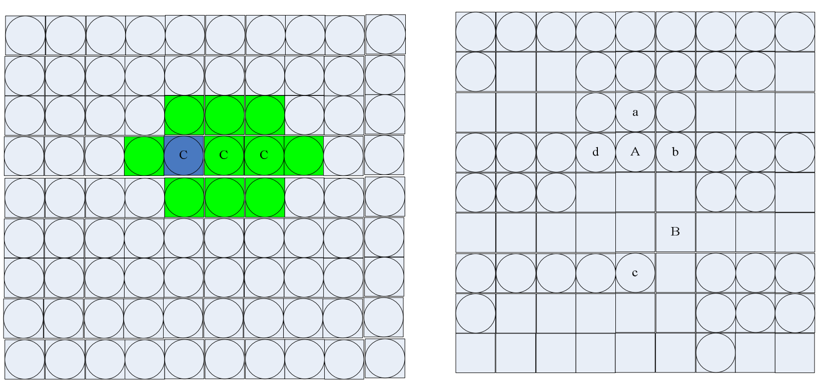


Fig. 3. The instruction of the Hill Climbing Local Search

In our implementation, after we selected the first n cells using the proposed methods in Clone and Mutation Kernel, the clone and mutation would start. In Memory Cells Local Search Kernel of our algorithm, we follow it by the implementation of eliminating those cells whose affinity is inferior to the given threshold.

3.5 Clone Suppression and Network Suppression

In the two suppression processes (clone suppression and network suppression), we need to calculate the affinity between the network cells. Based on the data structure we proposed, the calculation of Ab to all other Ab started. In this progress, we just need to judge if it satisfied the certain threshold and needn't save the calculation result. If it is inferior to threshold, the Is_Deleted flag is set to one, and then stop the calculation with the remaining cells, which means that it has been deleted from the network. The time complexity of this method is o (n), because one Ab need to compute with all other network cells once. We provide the flowchart in Fig.4, which has some detailed design as follows. The program judged the Is_Deleted at first to avoid the unnecessary calculation. If the Is_Deleted attribute has been set to one, this calculation will be canceled. The setting of this flag uses the atomic operation, that is to say, the data can be operated only by one thread at a time. Therefore, this mechanism also insures that we will not delete both the two Abs whose affinity between them is below the threshold.

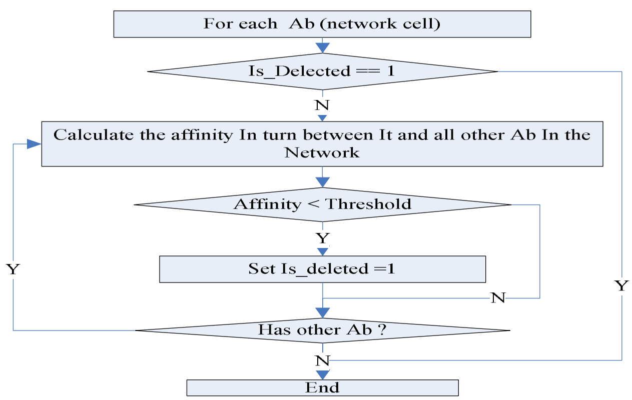


Fig. 4. The Flowchart of the Suppression

4 Experimental Results

A PC was used for execution and performance analysis: AIUS with Intel CPU i5 480 2.63 GHz processor, NVIDIA GeForce GT 425M. The applications presented in this paper were implemented on it. The CUDA timer functions are used to compute the time consumed by each operation. The speedup was evaluated in two experiments. The objective of first experiment is to test the speedup of different scales of records. And the second experiment shows that whether different number of antibodies influence the speedup.

In the first experiment, the data sets had 1000, 10,000, 100,000 records respectively, and each record set has three subsets, in which the record has two, four and eight attributes respectively. The antibody of our network is set to be 512 threads in one grid. From the Fig.5, we can see speedups for different number of clusters respectively of our result at the given number of threads. This speedup we got is due to that all the network cells parallel response with the antigen. From the Fig.5, we can see that the speedup increased with the data record size increased. However, due to that the antigens are present to the network serially, the speedup is not increased the same as the multiple of the data record increased. Also, we can see from the Fig.5 that the speedup is increased as the attributes increased, which is because that the affinity calculation processes are all modified to be parallel.

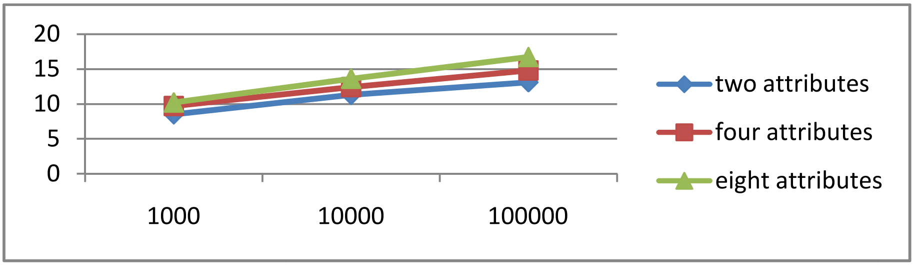


Fig. 5. Speedup with different size of data records

In the second experiment, there have 10000 records, and also have two, four, eight attributes respectively. However, the network size is given to 64, 256, 512, and 1024. As showed in Fig.6, all of them are tested under 1,000 records. With more network cells, more threads are started. However, speedup went up not so obviously, that is to say, the increase of network cells has little influence on speedup. Which is because that after the first selection process of network, the following steps are calculated on the first n cells selected. Therefore, there are no much differences achieved when change the network cells size.

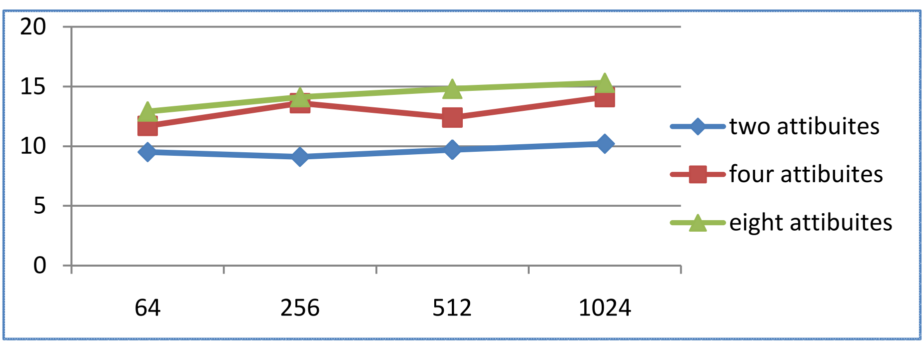


Fig. 6. Speedup with different network cells size

5 Conclusion

In this paper, parallelization of immune network based on the structure of GPU was investigated to improve the efficiency and scalability of AIN for data clustering. Firstly, we analyzed the time complexity of the aiNet algorithm and found that some parallel work could be done to improve the efficiency. We implemented our parallel immune network algorithm on GPU. The three kernels we had designed in section 3 executed on GPU to do the main task of data clustering. During the whole process, some skills were designed because we tried to make sure that the thread executed frequently to get higher speedup. For example, we combined the random numbers with the hill climbing algorithm as our local search method for the first n higher affinity cells; we carried out the network suppression in parallel introduced in section 3.6; we generated enough random numbers before the iteration process to make our algorithm high efficiency. The experiment result shows that the algorithm proposed have achieved about 10 times speedup than the original algorithm with the parallel technology.

It is noticed that the flow of the algorithm is relative complexity. Therefore, in the future work, we will try to optimize memory allocation and simplify the algorithm to make it more proper for practical using.

Acknowledgments. The research work described in this paper was fully supported by the grants from the National Natural Science Foundation of China (Project No. 90820010, 60911130513). Prof. Qian Yin are the author to whom all correspondences should be addressed.

References

1. De Castro, L.N., Von Zuben, F.J.: An evolutionary immune network for data clustering. In: Proc. Sixth Brazilian Symp. Neural Networks, pp. 84–89 (2000)
2. Timmis, J.: Artificial Immune Systems: A Novel Data Analysis Technique Inspired by the Immune Network Theory. Ph.D. Dissertation, Department of Computer Science, University of Wales (2000)
3. Thomas, S., Jonathan, T., Claudia, E.: A Comparative Study of Real-Valued Negative Selection to Statistical Anomaly Detection Techniques. In: Jacob, C., Pilat, M.L., Bentley, P.J., Timmis, J.I. (eds.) ICARIS 2005. LNCS, vol. 3627, pp. 262–275. Springer, Heidelberg (2005)
4. De Castro, L.N., Timmis, J.: An artificial immune network for multimodal function optimization. In: Proc. Congr. Evol. Comput., pp. 699–704 (2002)
5. Everitt, B., Landau, S., Leese, M.: Cluster Analysis. In: Everitt, B. (ed.), 4th edn. Hodder Arnold, London (2001)
6. Gao, X.B.: Fuzzy Clustering Analysis and Application. In: Gao, X.B. (ed.), pp. 49–160. Xidian University Press, Xian (2004) (in chinese)
7. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Fransisco (2000)
8. Jerne, N.K.: Towards a Network Theory of the Immune System. Ann. Immunol. 125C, 373–389 (1974)
9. Burnet, F.M.: Clonal Selection and After. In: Bell, G.I., Perelson, A.S., Pimbley Jr., G.H. (eds.) Theoretical Immunology, pp. 63–85. Marcel Dekker Inc., New York (1978)
10. Ada, G.L., Nossal, G.J.V.: The clonal selection theory. Scient. Am. 257, 50–57 (1987)
11. Hong, P., Wang, M., Lai, C.H.: Design of parallel algorithms for fractal video compression. Int. J. of Comput. Math. 84, 193–202 (2007)

12. Souravlas, S., Roumeliotis, M.: On further reducing the cost of parallel pipelined message broadcasts. Int. J. Comput. Math. 83, 273–286 (2006)
13. Sukumar, M., Madhumangal, P., Tapan, K.P.: Optimal sequential and parallel algorithms to compute a Steinertree on permutation graphs. Int. J. Comput. Math. 80, 939–945 (2003)
14. Watkins, A., Bi, X., Phadke, A.: Parallelizing an immune-inspired algorithm for efficient pattern recognition. In: Intelligent Engineering Systems through Artificial Neural Networks: Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Complex Systems and Artificial Life, vol. 13, pp. 225–230. ASME Press, New York (2003)
15. Watkins, A., Timmis, J.: Exploiting parallelism inherent in AIRS, an artificial immune classifier. In: Nicosia, G., Cutello, V., Bentley, P.J., Timmis, J. (eds.) ICARIS 2004. LNCS, vol. 3239, pp. 427–438. Springer, Heidelberg (2004)
16. Jianming, L., Lihua, Z., Linlin, L.: A Parallel Immune Algorithm Based on Fine-grained Model with GPU-Acceleration. In: 2009 Fourth International Conference on Innovative Computing, Information and Control (2009)

A Detection Method of Basic Mouth Shapes from Japanese Utterance Images

Tsuyoshi Miyazaki[1], Toyoshiro Nakashima[2], and Naohiro Ishii[3]

[1] Department of Information and Computer Sciences, Kanagawa Institute of Technology, 1030 Shimo-ogino, Atsugi, Kanagawa, Japan
[2] School of Culture-Information Studies, Sugiyama Jogakuen University, 17-3 Hoshigaoka-motomachi, Chikusa, Nagoya, Aichi, Japan
[3] Department of Information Science, Aichi Institute of Technology, 1247 Yachigusa, Yakusa, Toyota, Aichi, Japan
miyazaki@ic.kanagawa-it.ac.jp, nakasima@ci.sugiyama-u.ac.jp, ishii@aitech.ac.jp

Abstract. Some distinctive mouth shapes are formed when Japanese words and phrases are uttered. Because people who have acquired a skill of Japanese lip-reading know these characteristics, they can read lips movement. To realize the machine lip-reading, we propose a method which detects the distinctive mouth shapes from Japanese-speaking images based on their techniques. We define six mouth shapes as the distinctive mouth shapes, and the mouth shape images are used as template images. To detect the mouth shapes in utterance images, template matching is applied. Waveforms of similarity which are calculated by the template matching show some characteristic forms. Thus, we detect the mouth shapes from the waveforms. We carry out some experiments using Japanese words, and confirm effectiveness of the proposed method from the results.

Keywords: Lip-reading, Japanese phoneme, Template matching.

1 Introduction

Lip-reading is one of the important communication means for hearing-impaired people. In the lip-reading, a content of utterance is understood by movement of speaker's lips, etc. In recent years, some researches for realizing the lip-reading with information processing technologies have been pursued. This is called "machine lip-reading". It is used as a complementary technology to improve speech recognition, and the research of lip-reading for supporting communication with the hearing-impaired people is being pursued.

Generally, in the machine lip-reading, several images (frame images) of the lips region are taken using a camera or similar devices during a speaker is uttering. Digital image processing is performed on the frame images, and features about changes in mouth shape and movement of lips are calculated during an utterance. For example, a method which uses optical flow generates the features from velocity vectors[1]. A method which uses mouth shape changes generates the features from aspect ratio of

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 608–617, 2011.
© Springer-Verlag Berlin Heidelberg 2011

the lips region[2,3]. In addition, a method which uses images of the lips region generates the features from the images by template matching[4].

On the other hand, we found that people who have acquired a skill of lip-reading ("lip-reading skill holder" is used hereafter) stare at the mouth shape of speakers when reading lips. In addition, some distinctive mouth shapes are formed intermittently when uttering Japanese phones[1].

Thus, as the first step for realizing the machine lip-reading by modeling the lip-reading skill holder, we proposed a method in which knowledge of the lip-reading skill holders is logically materialized and the distinctive mouth shapes are processed using computers[5]. In this proposal, we defined six mouth shapes as "Basic Mouth Shape" (BaMS), and they are /a/, /i/, /u/, /e/, /o/ and closed mouth. In addition, there are some specific phones in which the formed mouth shape is different from the mouth shape of the vowel. We defined these mouth shapes of the specific phones as "Beginning Mouth Shape" (BeMS). For example, a closed mouth which is formed when we utter "*ma*" or "*pa*" is one of the BeMS. We defined the mouth shapes that are same as the mouth shapes of the vowel as "End Mouth Shape" (EMS).

After that, we defined some codes for each mouth shape of the BaMS as "Mouth Shape Code" (hereinafter called MS Code). We also proposed an expression method of the BaMS using the MS Code. Because the BaMS are formed sequentially when we utter arbitrary words, we defined an expression of the sequence of the BaMS using the MS Code as "Mouth Shapes Sequence Code" (MSSC). Furthermore, we defined the mouth shape patterns of all Japanese phones using the MS Code. We call the mouth shape patterns of the MS Code "Phone Code". As the result, it is possible to convert words to the MSSC[5].

We consider that it is possible to read lips if the BaMS are detected from Japanese-speaking images. We propose a detection method of the BaMS from Japanese-speaking images. To detect them, we adopt template matching which assumes the BaMS to the template images.

2 Detection of the Basic Mouth Shape

As for the movement of the mouth shape, it is a repetition of changes from a certain BaMS to another BaMS when Japanese is uttered[6]. Similarity for each BaMS is calculated by the template matching. On the waveforms of the similarity, a partial waveform is flat in the term in which the EMS is formed, and a partial waveform forms convex in case of the BeMS[7]. These characteristics of the waveform are utilized to detect the BaMS. The BaMS is defined as (1), and each symbol expresses mouth shape /a/, /i/, /u/, /e/, /o/ and closed mouth, respectively. The BeMS is defined as (2), and the EMS is defined as (3)[5].

$$BaMS = \{A, I, U, E, O, X\} \tag{1}$$

$$BeMS = \{I, U, X\} \tag{2}$$

$$EMS = \{A, I, U, E, O, X\} \tag{3}$$

[1] The length of voice equivalent of one short syllable is called "mora", and the voice is called "phone".

2.1 Detection Method of the End Mouth Shape

As mentioned above, the partial waveform is flat during the EMS is formed. Here, the similarity of the n-th frame and the mouth shape $m(\in EMS)$ is expressed as $R(m,n)$. Therefore, if (4) is satisfied for $\forall m$, the n-th frame is assumed "EMS frame" (TH is a threshold of the similarity). For all frames speaking Japanese, it is determined whether each frame is the ESM frame or not.. Accordingly, it is able to divide an utterance term between the "EMS terms" and the others, and we call the latter term "BeMS term".

$$\left|\Delta R(m,n)\right| \le TH \tag{4}$$

$$\Delta R(m,n) = R(m,n) - R(m,n-1) \tag{5}$$

2.2 Detection Method of the Beginning Mouth Shape

The BeMS are not mouth shapes formed by all means unlike the EMS[2]. Therefore an idea is necessary to detect the BeMS. We know that a partical convex waveform is formed when the BeMS is formed from our previous study[7], and the mouth shapes are detected using this characteristic.

At first, each similarity is analyzed in the BeMS term. In a term $n_a \le n \le n_b \left(n_a < n_b\right)$ (see Fig. 1) in which (6) is satisfied, if (7) or $n_b - n_a > N_F$ is satisfied, the waveform in the term is considered as an upward slant to the right. Here, D and N_F are thresholds of the similarity and number of the frames, respectively.

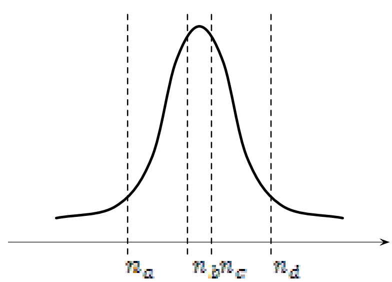


Fig. 1. The convex waveform which is formed when the BeMS is formed

$$\Delta R(m,n) > TH \tag{6}$$

$$\sum_n \left|\Delta R(m,n)\right| \ge D \tag{7}$$

[2] A phone which is uttered without the BeMS is called "Simple Mouth Shape Phone", and a phone which is uttered with the BeMS is called "Couple Mouth Shapes Phone".

Next, in a term $n_c \le n \le n_d \left(n_b < n_c < n_d\right)$ in which (8) is satisfied, if (7) or $n_d - n_c > N_F$ is satisfied, the waveform in the term is considered as a downward-sloping.

$$\Delta R(m,n) < -TH \tag{8}$$

Finaly, if (4) and $n_c - n_b \le N_P$ are satisfied in a term $n_b \le n \le n_c - 1$, it is considered to be formed a BeMS between the n_a-th frame and the n_d-th frame from the waveform. Here, N_P is a threshold of number of the frames.

These processes are applied for each mouth shape of the BeMS, and peak value $R_P(m)$ is calculated by (9). The $R_P(m)$ is used for detection of the BeMS.

$$R_P(m) = \max\left(R(m,n_b), R(m,n_b+1), \cdots, R(m,n_c-1)\right) \tag{9}$$

3 Experiment for the Detection of the Basic Mouth Shape

To evaluate the method we proposed, we carry out some experiments to detect the BeMS and the EMS. The experiment contents are as follows:

1. Detecting the EMS from words that are composed by the Simple Mouth Shape Phone.
2. Detecting the BeMS from simple words that are composed by the Couple Mouth Shapes Phone.
3. Detecting the BaMS from words

Configuration of equipment carrying out the experiments is shown in Fig. 2. Images around mouth of a subject are taken with a digital video camera (hereinafter called DV camera), and the images are transferred to a PC.

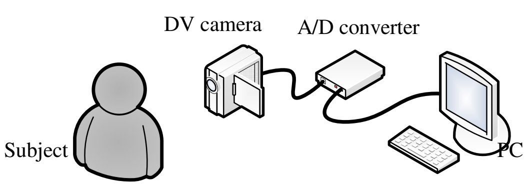


Fig. 2. Configuration of the experiments

During the experiments, distance between the DV camera and the mouth of subject are kept constantly, and the subject utters without moving head. Because the size of mouth is kept constantly and the horizontal and vertical movement of mouth is minimized. Before detecting the mouth shapes, the BaMS images which are used for the template are taken (see Fig. 3). In this experiment, the size of the frame images is assumed 640×480 pixels, and the size of the template images is trimmed to 495×375

pixels. When the template matching is performed, the images are converted into gray-scale images, and normalized cross-correlation is applied. The thresholds of constant to detect the BeMS and EMS are shown in Table 1. The first mouth shape and the last mouth shape of an utterance are assumed the closed mouth, and these two mouth shapes are not targeted for the detection.

Fig. 3. Template images

Table 1. The thresholds of constant

TH	D	N_F	N_P
0.02	0.06	3	4

3.1 Detection of the End Mouth Shape

The words used in this experiment are shown in Table 2. Each word is uttered five times. From the mouth shapes for which the averages of the similarity are greater than 0.6, up to two EMS having the highest and the second value are selected. In the selected BaMS, it is decided whether the EMS which should be detected is included or not.

Table 2. Test words and its MSSC to detect the EMS

#	Words (Japanese syllabary)	MSSC
1	ATSUGI	-A-U-I
2	KOPPU	-O-X-U
3	KOKESHI	-O-E-I
4	ENPITSU	-E-X-I-U
5	KONPAIRU	-O-X-A-I-U

The detection rates of EMS are shown in Table 3. The detection rates of each EMS which is included in the MSSC are shown in the column of “Detection rates”, and the column of “As the highest similarity” shows the detection rates of the EMS which is detected as the highest average similarity. As a whole, the average detection rate of the EMS which was included in the two selected BaMS was 98.9%, and the rate which was detected the EMS as the highest similarity was 95.6%. We consider that these results are satisfactory. An example of the waveforms of the similarity is shown in Fig. 4. The horizontal axis of the chart shows the frame number, and the vertical axis shows the similarity R. The range of R is between -1.0 and 1.0, and 1.0 shows the best similarity. In addition, vertical lines in the chart separate the BeMS term and the EMS term. The ESM terms are shown with an arrow, and the previous terms are the BeMS term.

Table 3. Detection rates of the EMS

#	Words	Detection rates	As the highest similarity
1	ATSUGI	100.0%	100.0%
2	KOPPU	100.0%	100.0%
3	KOKESHI	100.0%	93.3%
4	ENPITSU	100.0%	95.0%
5	KONPAIRU	96.0%	92.0%
	Average	98.9%	95.6%

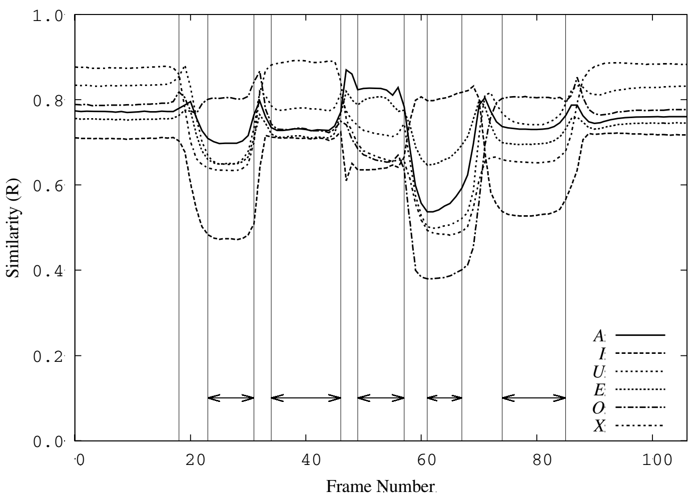


Fig. 4. Waveforms of the similarity about the test word "KONPAIRU"

3.2 Detection of the Beginning Mouth Shape

We choose simple words that are composed with two phones because detection of the BeMS is the purpose in this experiment. The test words are shown in Table 4. Same as the experiment of the EMS, the BeMS are detected from five utterances.

Table 4. Test words and its MSSC to detect the BeMS

#	Words (Japanese syllabary)	MSSC	#	Words (Japanese syllabary)	MSSC
1	ASA	-AIA	6	ASE	-AIE
2	NIWA	-IUA	7	UME	-UXE
3	UMA	-UXA	8	ASO	-AUO
4	KAMI	-AXI	9	IMO	-IXO
5	GAMU	-AXU			

The detection rates of the BeMS are shown in Table 5. As a whole, the detection rate of the BeMS was 93.3%, and the rate which was detected the BeMS as the highest peak value R_P was 88.9% (the column of "As the highest similarity"). We also consider that these results are satisfactory. An example of the chart is shown in Fig. 5. In the BaMS term which is shown with an arrow, the waveform of X forms convex. As a result, X is detected.

Table 5. Detection rates of the BeMS

#	Words	Detection rates	As the highest similarity
1	ASA	100.0%	100.0%
2	NIWA	100.0%	100.0%
3	UMA	100.0%	80.0%
4	KAMI	80.0%	80.0%
5	GAMU	100.0%	100.0%
6	ASE	100.0%	100.0%
7	UME	100.0%	80.0%
8	ASO	60.0%	60.0%
9	IMO	100.0%	100.0%
	Average	93.3%	88.9%

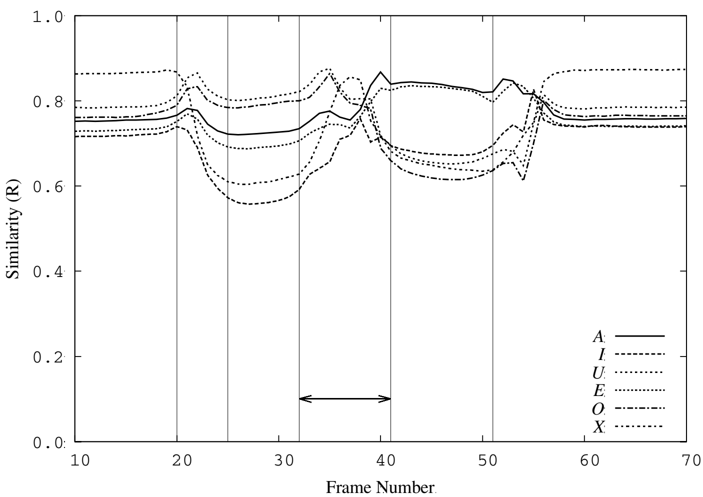


Fig. 5. Waveforms of the similarity about the test word "UMA"

3.3 Detection of the Basic Mouth Shape

The test words are shown in Table 6. Same as the previous experiments, the BaMS are detected from five utterances. Each detection rate is shown in Table 7. "Detection rates" shows the detection including the BeMS and the EMS, and the column of "False detection rates" shows the detection rate of the wrong BeMS which is not formed in the term. The detection rates were uneven, but the average detection rate of the BaMS was 75.6%as a whole. The average detection rate of the BeMS was 67.3%, and the average detection rate of the EMS was 79.4%. These results are lower than the previous results. On the other hand, the false detection rate was 4.0%. Besides, the detection rate was 0.0% in some words. An example of the chart is shown in Fig. 6.

Table 6. Test words and its MSSC to detect the BaMS and the EMS

#	Words (Japanese syllabary)	MSSC
1	KATATSUMURI	-AIA-UXU-I
2	KAWAKUDARI	-AUA-UIA-I
3	KAMISHIBAI	-AXIXA-I
4	ASESUMENTO	-AIE-UXE-IUO
5	SUPOTTORAITO	-UXO-U-OIA-IUO

Table 7. Detection rates of the BaMS

#	Words	Detection rates	Detection rates of BeMS	Detection rates of EMS	False detection rates
1	KATATSUMURI	68.6%	100.0%	56.0%	13.3%
2	KAWAKUDARI	85.7%	50.0%	100.0%	0.0%
3	KAMISHIBAI	100.0%	100.0%	100.0%	0.0%
4	ASESUMENTO	68.9%	53.3%	76.7%	6.7%
5	SUPOTTORAITO	55.0%	33.3%	64.3%	0.0%
	Average	75.6%	67.3%	79.4%	4.0%

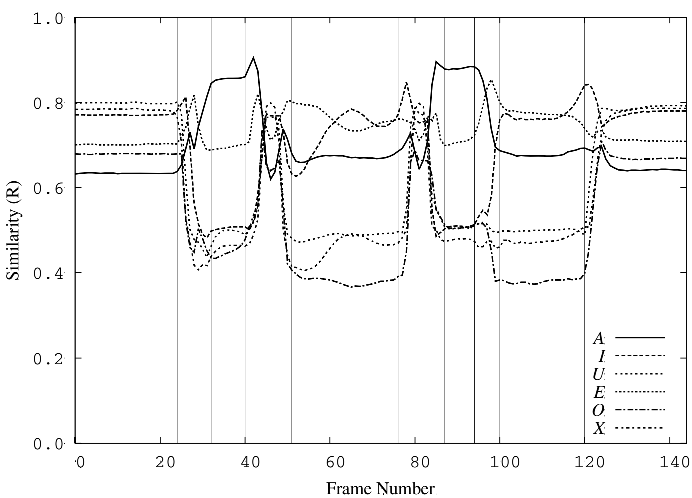


Fig. 6. Waveforms of the similarity about the test word “KAMISHIBAI”

We consider the cause that the detection rates of the BeMS and the EMS lowered. The underscored MS Code of the BaMS which have low detection rates are shown in Table 8. As a result, it may be said that the detection rates lower about I and U. However, all I and U are not low. As a result of analysis, we recognized two cases.

Table 8. The BaMS of low detection rate

#	Words	MSSC
1	KATATSUMURI	-AIA-UXU-I
2	KAWAKUDARI	-AUA-UIA-I
3	KAMISHIBAI	-AXIXA-I
4	ASESUMENTO	-AIE-UXE-IUO
5	SUPOTTORAITO	-UXO-U-OIA-IUO

Firstly, there are two mouth shape images about only U. One is an image in which teeth appear, the other is an image in which teeth do not appear. When we utter phones such as “SU”, “TSU” or “NU”, the teeth appear in the image. For this reason, we consider that the detection ratio falls.

Secondly, in the case the mouth shape of the BeMS is similar to the next mouth shape of the EMS, it is difficult to detect the BeMS. For example, it is difficult to detect U of the BeMS such as “SO” or “TO”, because the mouth shapes of U and O are similar. So, we have to consider the solution to these problems.

4 Conclusion

In this paper, we proposed a method which detects the BaMS from Japanese-speaking images. We adopt template matching to detect the BaMS, and the similarities were calculated. From the previous study, we confirmed that the waveforms of the similarity formed unique form when the BeMS and the EMS are formed. So, we paid attention to these characteristics and were able to detect the BaMS. We confirmed that the proposal method was effective from the experiments. At the same time, we also recognized some mouth shape patterns that were difficult to detect the BaMS. We have to review the solution to the problems, and it is necessary to improve the detection accuracy in future.

References

1. Kenji, M., Pentland, A.: Automatic Lip-reading by Optical-flow Analysis. The Transactions of the Institute of Electronics. Information and Communication Engineers J73-D-II(6), 796–803 (1990) (in Japanese)
2. Yasuyuki, N., Moritoshi, A.: Lipreading Method Using Color Extraction Method and Eigenspace Technique. The Transactions of the Institute of Electronics, Information and Communication Engineers J85-D-II(12), 1813–1822 (2002) (in Japanese)

3. Takeshi, S., Mitsugu, H., Ryosuke, K.: Analysis of Features for Efficient Japanese Vowel Recognition. The IEICE Transactions on Information and Systems E90-D(11), 1889–1891 (2007)
4. Kimiyasu, K., Keiichi, U.: An Uttered Word Recognition Using Lip Image Information. The Transactions of the Institute of Electronics, Information and Communication Engineers J76-D-II(3), 812–814 (1993) (in Japanese)
5. Tsuyoshi, M., Toyoshiro, N.: The Codification of Distinctive Mouth Shapes and the Expression Method of Data Concerning Changes in Mouth Shape when Uttering Japanese. IEEJ Transactions on Electronics, Information and Systems 129(12), 2108–2114 (2009) (in Japanese)
6. Tsuyoshi, M., Toyoshiro, N., Naohiro, I.: Evaluation for an Automatic Generation of Lips Movement Images Based on Mouth Shapes Sequence Code in Japanese Pronunciation. In: Proc. of Japan-Cambodia Joint Symposium on Information Systems and Communication Technology 2011 (JCAICT 2011) pp. 89–92 (2011)
7. Tsuyoshi, M., Toyoshiro, N.: A Deriving Method of Distinctive Mouth Shapes for Lip-reading. In: Multimedia, Distributed, Cooperative, and Mobile Symposium, pp. 1544–1549 (2009) (in Japanese)

Improving the Usability of Hierarchical Representations for Interactively Labeling Large Image Data Sets

Julia Moehrmann[1], Stefan Bernstein[2], Thomas Schlegel[3], Günter Werner[2], and Gunther Heidemann[1]

[1] Intelligent Systems Group, University of Stuttgart, Universitätsstr. 38, 70569 Stuttgart, Germany
{moehrmann,heidemann}@vis.uni-stuttgart.de
[2] University of Apllied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany
{sbernste,gwerner}@hs-mittweida.de
[3] Softwareentwicklung ubiquitärer Systeme, TU Dresden, Nöthnitzer Straße 46, 01187 Dresden, Germany
thomas.schlegel@tu-dresden.de

Abstract. Image recognition systems require large image data sets for the training process. The annotation of such data sets through users requires a lot of time and effort, and thereby presents the bottleneck in the development of recognition systems. In order to simplify the creation of image recognition systems it is necessary to develop interaction concepts for optimizing the usability of labeling systems. Semi-automatic approaches are capable of solving the labeling task by clustering the image data unsupervised and presenting this ordered set to a user for manual labeling. A labeling interface based on self-organizing maps (SOM) was developed and its usability was investigated in an extensive user study with 24 participants. The evaluation showed that SOM-based visualizations are suitable for speeding up the labeling process and simplifying the task for users. Based on the results of the user study, further concepts were developed to improve the usability.

Keywords: Self-organizing map, SOM, user study, image labeling, ground truth data.

1 Introduction

The importance of image recognition systems increases with the ubiquity of webcams and camera phones. However, the extensive development of image recognition systems for non-industrial purposes fails due to time and cost. One important factor is the labeling (or annotation) of the training images. Labeling the image data is necessary for building a ground truth data set on which the classifier for a specific recognition system can be trained. Since correctness in the ground truth data is crucial for the recognition rate, labeling has to be performed manually. Since image recognition systems need large amounts of images to cover the appearance space in all its details (variances in lighting, rotation, occlusion ...) it takes a lot of effort to assign correct labels to all images.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 618–627, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Conquering the complexity of labeling large image data sets is possible by using a semi-automatic approach. Here, the image labeling is not performed manually on the individual images, but instead features are used for clustering the image data unsupervised. A resulting cluster represents similar images and can be labeled by the user in one go. The labeling itself is still performed manually, since an unsupervised clustering may result in heterogeneous clusters, i.e. clusters containing errors. However, the necessary interaction can be dramatically reduced.

We implemented a semi-automatic approach, using a self-organizing map (SOM) to cluster image data. Based on this system we developed concepts for investigating and labeling the clustered image data set. For this purpose, we provide a sophisticated user interface which allows for an easy and intuitive labeling of images and gives a good overview of the complete data set, and especially of the parts that still need to be labeled. We conducted a user study with 24 participants where the users were asked to annotate three different data sets. The goal of the user study was to find out whether the developed concepts simplify the interactive labeling task for the user. Based on the study results further concepts were developed to optimize the user experience.

2 Related Work

The creation of ground truth data sets has been neglected as a research topic for a long time. The idea of human computation has led to the development of a few tools for the creation of general purpose ground truth data sets, like the ESP Game [1], Peekaboom [2], and systems by Ho et al. [3] and Seneviratne et al. [4]. People participate in these games because they are entertaining, not because of the actual benefit. Russell et al. provide a web-based interface (LabelMe) for labeling images [5]. The incentive for contributing is the data itself. Similarly, Yao et al. [6] provide an annotation tool for the creation of a large ground truth data set. A semi-supervised labeling process is realized through a hierarchical segmentation of images. All of these approaches aim at the creation of general purpose data sets. They are, however, not applicable to the task of labeling specific data sets which include less variance.

In the area of content based image retrieval, Koskela et al. [7] developed a system for the automated semantic annotation of images based on the PicSOM system [8]. An existing ground truth data set is used to annotate other images in the data set. Heidemann et al. developed VALT (Visualization and Labeling Toolkit) [9] for labeling images clustered with a self-organizing map. Category labels can be assigned to nodes or individual images. However, the system was intended for AR scenarios.

A variety of systems and techniques for browsing large digital photo collections have been proposed. Most photo browsing applications use hierarchical representations. Among the best known applications are PhotoMesa [10], PhotoTOC [11], and CAT [12]. All three cluster images in order to provide a suitable interface for browsing. Both PhotoMesa and CAT use hierarchical structures and representative images for clusters in each hierarchy.

Despite these developments the problem of labeling large specialized data sets has not yet been solved. Therefore concepts are necessary to handle the complexity of large data sets without the possibilities of classical human computation approaches.

3 Data Preprocessing and Self-Organizing Map

For conducting our user study we extracted color histogram features as the input for clustering. However, any other type of features may be used. The feature vectors are processed by a standard SOM as proposed by Kohonen [13]. The SOM Toolbox for Matlab[14] is used for SOM calculation.

A SOM is a type of unsupervised neural network which projects high-dimensional input data to a low-dimensional (usually 2D) map (or grid). The input to a SOM are *n*-dimensional feature vectors. Each unit of the SOM grid is associated with one weight vector of the same size as the input vectors. The SOM is trained by iteratively calculating the best matching unit (BMU), i.e. the unit with the smallest euclidean distance to an input vector. The weight vectors of the BMU and its neighboring units are adapted towards the input vector. The shape and size of the neighborhood, as well as the learning rate, is controlled with a neighborhood function, usually a Gaussian.

An important result of this training process is that SOMs are topology preserving, i.e. data vectors that are close in the input space will be close in the resulting map. This aspect makes SOMs extremely valuable in data exploration tasks.

4 Interaction and Visualization Concepts

Figure 1a) displays the visualization of a trained SOM using a representative image for each node. A representative image is taken from the set of images which were assigned to this node. Additionally, we implemented a second visualization using the Unified Distance Matrix (U-Matrix) [15] (see Figure 1b)). Basically, the U-Matrix is a color-coded visualization of the mean distance to neighboring map units. The U-Matrix is widely used for visualizing clusters (dense areas) on trained SOMs. Therefore we wanted to investigate whether the color-coding is also suitable if image data is used. The arrangement of images is the same for both visualizations.

As mentioned, the quality of the ground truth data is crucial for training classifiers. Since the SOM clusters the data unsupervised, it is important that it is still possible for users to take a closer look at and label each image individually. Therefore, the developed user interface is zoomable and allows the user to view the SOM representation in both levels of detail (LOD), the map view and the detail view.

Figure 1a) shows the SOM where each unit is displayed by a representative image, c) shows the detail view of three selected units. In the detail view all images which were projected onto these map units are displayed. The utilization of more levels of detail were considered but not realized due to the increasing complexity and the inferior overview of the data set. In general more levels of hierarchy are preferable for image browsing tasks, however, if every single image needs to be processed by the user, a simple structure is desirable.

Although the SOM visualization itself has two LODs only, the interface provides a natural zooming behavior, since images move apart as the zoom factor is increased. However, the images are not scaled because larger images would require the user to draw wider selection areas.

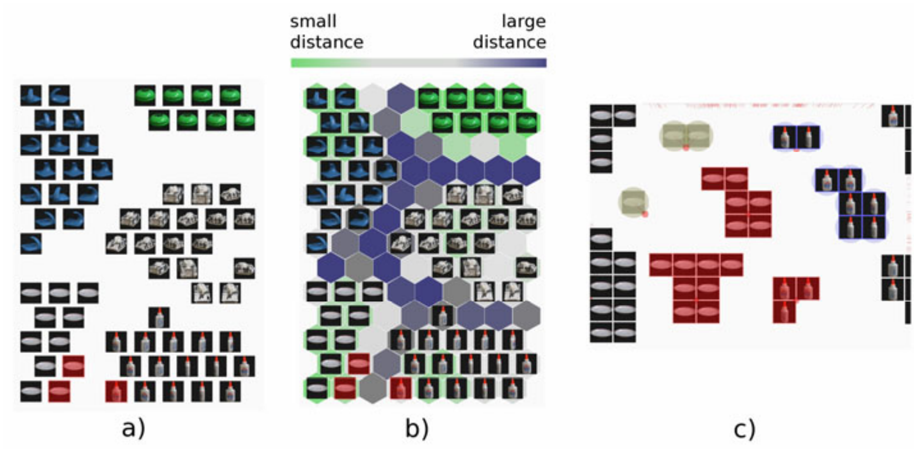


Fig. 1. SOM clustered visualization of COIL-100 subset [16]. a) shows the plain standard visualization of the map view, b) displays the color coded U-Matrix in the background on the hexagonal map. Selected map units are highlighted in red. c) shows the detail view of the area with the selected map units. Grayed out images are labeled, as well as images with blue boundaries. Subsidiary lines are displayed on upper boundary of c).

Inherent to a zoomable user interface for labeling large image data sets is the fact that only a subset of all images can be displayed at once. To simplify the navigation subsidiary lines were introduced to indicate the direction of remaining unlabeled images outside the current viewport (see Figure 1c)).

The user interface allows the direct selection of images, as well as the use of a rectangle and a free selection (lasso) tool. There is no difference in the selection mechanism of units or individual images in the detail view. If a node is selected and labeled, the label is assigned to all images attached to this node. However, in the detail view images may be selected and labeled individually, thus allowing the user to correct possible clustering errors.

Users can define labels in the interface. For each label a button with the name and a customized color is displayed in a separate panel. Labeled images are highlighted in the color of the assigned label and are drawn semitransparent as shown in Figure 1c) for plate images (gray). The images are displayed semitransparent because they are no longer selectable, thus simplifying the selection of the remaining images. In case a user wants to reassign or remove a label from an image, the visibility may be toggled and the images are again drawn opaque, though still with highlighted boundary, and can be selected. Toggling the visibility refers to all images assigned to the selected category label, as in Figure 1c) where all images assigned to *bottle* are opaque, i.e. selectable. This function was also intended to simplify the examination of all images that were assigned to a certain category. If only the images belonging to a certain category are displayed opaque, it is likely that incorrectly assigned images stand out visually and can be detected more easily in a final check. An option to hide all labeled images was included in the interface to further reduce visual clutter.

5 Evaluation

The concept of an unsupervised clustering with a SOM as the basis for a zoomable user interface is quite intuitive and promising. Nevertheless, unsupervised learning algorithms provide no guarantee that the resulting clusters agree with the users' expectations. A nice clustering in the map view might give users a false sense of security regarding the correctness of the unsupervised arrangement. Incorrectly clustered images may be missed by users and lead to inaccurate ground truth data sets. Additionally, it is important to investigate whether our proposed concepts are actually suitable for simplifying and accelerating the labeling task. For comparison, a third visualization was realized, which will be referred to as "unsorted representation".

The study comprises different labeling tasks. Table 1 shows a summary with the data sets used. In data sets 1 and 2 the number of images for individual objects varies from 1 to 72 and was chosen randomly. Data set 3 is a real data set with webcam images showing windows.

The sequence of the tasks was the same for all participants. The tests within each task, i.e. the different visualizations, were randomized to compensate for learning effects.

Table 1. Summary of labeling tasks and the data sets used. The same visualizations were tested in T2 and T3. The discussion of hypothesis 1-3 can be found in Section 5.2.

Task	Data set	#Images	#Labels	Visualizations	Hypotheses
T1	COIL-100	432	5	plain SOM	H3
T2	COIL-100	857	18	Unsorted, SOM, U-Matrix	H1, H2
T3	Webcam/ Windows	666	2	Unsorted, SOM, U-Matrix	H1, H2

5.1 Study Method and Procedure

The user study we conducted used a within-subjects design. In this study, we investigated the usability of the SOM based labeling interface, both with and without the U-Matrix visualization. Since no standard approach exists for labeling specific image data sets the visualization denoted as unsorted representation was used as baseline. The behavior of the user interface is exactly the same as for the SOM-based visualizations except that there is only one level of hierarchy. Instead all images are visible in one plane, exactly in a quadratic grid aligned by rows. Images were not randomized but instead displayed according to their filenames. The COIL set is special due to the labels being embedded in the filenames. Labeling of this data could be simplified using an explorer like interface. However, the COIL data is well suited for presenting the functionality of the system. In contrast, the data set in T3 did not include such a structure.

24 persons participated in the study (21 male, 3 female). 20 participants are computer scientists, four participants are students. The average age was 29 years (min. 16, max. 50). Ten participants were familiar with image recognition. All participants reported normal or corrected-to-normal vision. Two participants suffer from dyschromatopsia. Their results were within range of the other results and are therefore included in all calculations.

The study began with every participant answering personal statistical questions. The participants were then introduced to the topic of the study in a 15min tutorial. The tutorial included the SOM visualizations and the unsorted representation. Participants were asked to use the GUI in order to get familiar with the controls.

Completion times for individual tests and the assigned labels were recorded. After each test (each visualization), participants were asked to judge the simplicity (Q1) and speed (Q2) on a 7-point Likert scale (1 -- strongly disagree, 7 -- strongly agree). The last part of the evaluation consisted of general questions about the user interface.

Participants needed approximately 1.5 hours to complete the study. Labels were predefined for all tasks. In three cases participants interchanged labels for whole categories. These labels were manually corrected since they were caused by mere misunderstanding.

Table 2. *p*-values calculated using paired t-tests. Simplicity and speed were judged by participants on a 7-point Likert scale after each test. Average values are given for all parameters, standard deviation is given in brackets. Error rate calculated as η = #incorrectLabels / #images. Participants were asked for their favorite visualization after the completion of whole tasks.

	T2			T3		
	Unsorted	SOM	SOM w. U-Matrix	Unsorted	SOM	SOM w. U-Matrix
p	< 0.001			< 0.001		
duration (avg)	727s (237)	412s (146)	430s (131)	503s (147)	293s (144)	309s (121)
p	0.009			<0.001		
Q1 (avg)	4.31 (1.49)	5.17 (1.02)	5.57 (0.89)	3.35 (1.46)	4.95 (1.43)	5.05 (1.19)
p	<0.001			<0.001		
Q2 (avg)	2.7 (1.29)	5.35 (1.15)	5.74 (0.81)	2.8 (1.44)	5.25 (1.25)	5.05 (1.15)
η Error	0.0045	0.011	0.016	0.279	0.27	0.275
Favorite vis.	1	13	10	4	5	15

Experiment Setup. The tests were performed on two laptop computers. The first one has an Intel Core 2 Duo 2.8GHz with 4GB RAM, an NVIDIA GeForce 9600M GT, and a 15 inch display with 1440 x 900 pixels resolution. The second computer has an AMD Athlon II Dual Core M300 2.0 GHz with 4GB RAM, an ATI Mobility Radeon HD 4200, and a 15,6 inch display with 1366 x 768 pixels resolution. Custom laptops with standard display sizes were used to investigate the usability under usual conditions. The differences of the GPUs were regarded as negligible since only two-dimensional graphics are used.

Groups of 12 participants completed the study on each laptop. There was no statistically significant difference in the completion times or error rates of both groups. The following results, therefore, refer to all 24 participants.

5.2 Hypotheses

The evaluation was conducted to test a total of three hypotheses. The tests and the evaluation for each hypothesis are given individually in the following sections.

H1: Arranging images with a SOM simplifies the labeling task. This section investigates the unsorted representation and the plain SOM visualization. Considering the completion times in tasks 2 and 3 (see Table 2), the results are statistically significant. In both tasks the average time to complete the labeling using the SOM visualization was less than 60% of the time needed for the unsorted representation.

Additionally, the subjective preferences (Q1 and Q2) in both tasks clearly favor the SOM visualization. The large differences in the judgment of speed are most probably due to the fact that users did not have to look at every single image.

In T2 the SOM visualization exhibits a larger error rate than the unsorted representation. In the unsorted representation all images are displayed in a two-dimensional plane without any hierarchy. Every image is therefore immediately visible, which leads to only a few incorrectly labeled images. The SOM clustering contained a few nodes that were not correctly clustered. Several participants chose to trust the SOM clustering and neglected the detail view, while others did look at the detail view but missed some errors anyway. In T3, the SOM visualization exhibits a slightly lower error rate. This may be due to the SOM representation reaching a very good separation of closed and opened windows.

H2: Visualizing the U-Matrix helps identifying clusters. It is not possible to get statistically significant results for both SOM based visualizations due to their inherent similarity. Nevertheless, to give an overview of the results, Table 2 gives the average and standard deviation of the duration, as well as for the questions about simplicity and speed, and the relative error.

When asked for their favorite visualization at the end of T2 and T3 the U-Matrix version was mentioned 25 times (see Table 2). In the final interview all participants were asked whether they found the U-Matrix visualization helpful. 15 participants answered that they did not look at the colors or did not care about them, but found the hexagonal grid structure helpful for navigating inside the data set. The grid helped participants to estimate the current zoom level since hexagons scale as the zoom factor is increased. Three participants mentioned that they found the colors distracting but liked the hexagonal structure. Only four participants stated that they found the distance encoding at least partially helpful. Since images themselves are easy to grasp it is not surprising that the U-Matrix color is of assistance in specific cases only.

H3: Drawing labeled images semitransparent and making them non-selectable enhances visual clarity. In T1 participants labeled the data set twice: once labeled images were drawn semitransparent and were non-selectable, the other time labeled images were highlighted and drawn opaque. The two representations were described above in the context of toggling the visibility of labeled images. In this test the representations were used exclusively and could not be switched.

The clustering in this example included no errors and the separation of clusters was clearly visible. When asked about their impression 14 participants answered that they preferred the version with semitransparent and non-selectable images because it provided a better overview of remaining unlabeled images. It was mentioned 8 times that the selection of remaining images was simpler if labeled images were non-selectable. 6 participants felt there was no significant difference in using both versions.

These results imply that the semitransparent representation of labeled images provides a benefit. However, participants used the option to hide all labeled images in a majority of the following tests. We therefore conclude that the additional colors present visual clutter even if it is less strong.

5.3 Further Enhancements

Although participants clearly favored SOM based visualizations over the unsorted representation, the risk of a higher error rate cannot be underestimated. One major drawback is the representation of nodes through one random image. We therefore developed four alternative representations which allow users to detect clustering errors more easily. These representations are average images, eigenimages and subset images. Subset images were calculated from the first four eigenvectors for one representation. Another subset image is calculated using k-means clustering on each unit (k = 4) and combines images closest to the cluster centers. Examples are given in Figure 2.

Although all representations allow the detection of clustering errors, the average images were not appreciated by users. Representations based on eigenimages were experienced as least suitable. Although clustering errors of the SOM are well visible, they complicate the labeling of correctly clustered nodes, since the object cannot directly be identified (see Figure 2a-b)). The most intuitive and helpful representation proved to be the subset image based on k-means clustering. It displays the images in a very intuitive and familiar way and allows easy identification of incorrect clusters. By deploying these representations another disadvantage was eliminated. Users are now able to grasp their position inside the hierarchy, since it is obvious whether the map or detail view is displayed.

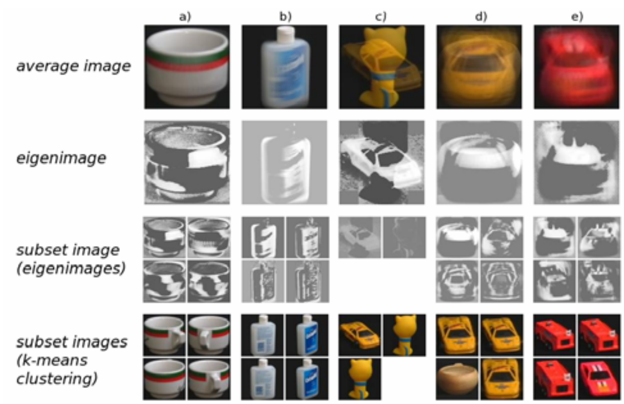


Fig. 2. Alternative representations for map units. Average images allow the detection of clustering errors, but cause a nauseating feeling, Eigenimages complicate the identification of the actual object. Subset images from k-means clustering seem intuitive and display four very different images belonging to this unit. Identical objects as in a) and b) can be grasped immediately by users.

Another improvement was the adaption of the U-Matrix color scheme. Initially a continuous red to blue gradient was used, which was impractical and too distracting. Instead, only very dense and very sparse areas are highlighted. Medium distances are displayed in neutral gray to reduce the amount of information transported via color.

6 Conclusion

We developed a graphical user interface concept for the interactive semi-automatic labeling of large image data sets. The interface uses a clustering calculated by a self-organized map (SOM) as the basis for a visualization, with or without the U-Matrix to transport additional information about the node distances.

The interface was developed to investigate concepts for optimizing the usability when labeling large image data sets. This was realized by drawing labeled images semitransparent and making them non-selectable. The zoomable interface provides an overview of the SOM and still lets the user explore all images belonging to one unit. To investigate whether the SOM based visualizations are applicable to real image labeling tasks a user study with 24 participants was conducted. Although the SOM visualizations yield a risk for a slightly higher error rate, the usability was preferred over labeling images sorted by filename. SOM visualizations achieved shorter completion times and were subjectively experienced as faster and easier to use. The drawback of the SOM visualizations is that they might yield a false sense of security regarding the unsupervised clustering. Therefore, alternative representations for node images were developed which reduce the risk of missing clustering errors. Additionally, the U-Matrix color coding was adapted to better fit user needs.

Future work will include research on whether existing visualizations used in SOM based data exploration, like hit histograms or component planes, can be of advantage in supporting the image labeling task. A major topic will also be the integration of user feedback for an iterative recalculation of the SOM.

We believe the developed concepts will become more important in the future as the need to annotate large image data sets increases. Nevertheless, it will be necessary to extend the proposed concepts in the future to further increase the usability of labeling systems, especially for very large data sets.

References

1. von Ahn, L., Dabbish, L.: Labeling Images with a Computer Game. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 319–326 (2004)
2. von Ahn, L., Riu, R., Blum, M.: Peekaboom: a Game for Locating Objects in Images. In: SIGCHI Conference on Human Factors in Computing Systems (CHI 2006), pp. 55–64 (2006)
3. Ho, C.-J., Chang, T.-H., Lee, J.-C., Hsu, J.Y.-J., Chen, K.T.: KissKissBan: a Competitive Human Computation Game for Image Annotation. In: ACM SIGKDD Workshop on Human Computation, pp. 11–14 (2009)
4. Seneviratne, L., Izquierdo, E.: An Interactive Framework for Image Annotation through Gaming. In: International Conference on Multimedia Information Retrieval, pp. 517–526 (2010)

5. Russel, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: LabelMe: A Database and Web-based tool for image annotation. Int. J. on Computer Vision 77(1), 157–173 (2008)
6. Yao, B., Yang, X., Zhu, S.-C.: Introduction to a Large-Scale general Purpose Ground Truth Database: Methodology, Annotation Tool and Benchmarks. In: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recgnition, pp. 169–183 (2007)
7. Koskela, M., Laaksonen, J.: Semantic Annotation of Image Groups with Self-Organizing Maps. In: Leow, W.-K., Lew, M., Chua, T.-S., Ma, W.-Y., Chaisorn, L., Bakker, E.M. (eds.) CIVR 2005. LNCS, vol. 3568, pp. 518–527. Springer, Heidelberg (2005)
8. Laaksonen, J., Koskela, M., Laakso, S., Oja, E.: PicSOM – Content-based Image Retrieval with Self-Organizing Maps. Pattern Recognition Letters
9. Heidemann, G., Saalbach, A., Ritter, H.: Semi-Automatic Acquisition and Labelling of Image Data using SOMs. In: European Symposium on Artificial Neural Networks, pp. 503–508 (2003)
10. Bederson, B.: PhotoMesa: a Zoomable Image Browser using Quantum Treemaps and Bubblemaps. In: ACM Symposium on User Interface Software and Technology, pp. 71–80 (2001)
11. Platt, J., Czerwinski, M., Field, B.: PhotoTOC: Automatic Clustering for Browsing Personal Photographs. In: Joint Conference of the International Conference on Information, Communications and Signal Processing, and the Pacific Rim Conference on Multimedia, vol. 1, pp. 6–10 (2003)
12. Gomi, A., Miyazaki, R., Itoh, T., Li, J.: CAT: a Hierarchical Image Browser using a Rectangle Packing Technique. In: International Conference on Information Visualisation, pp. 82–87 (2008)
13. Kohonen, T.: The Self-Organizing Map. Proceedings of the IEEE 78(9), 1464–1480 (1990)
14. Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J.: Self-Organizing Map in Matlab: the SOM toolbox. In: Matlab DSP Conference, pp. 35–40 (1999)
15. Ultsch, A., Sieman, H.P.: Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis. In: Proceedings of International Neural Networks Conference, pp. 305–308. Kluwer Academic Press, Dordrecht (1990)
16. Nene, S., Nayar, S., Murase, H.: Columbia Object Image Library (COIL-100). Technical report, Columbia University (1996)

The Similarity Index of Character Shape of Medicine Names Based on Character Shape Similarity (II)

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

[1] Shibaura Institute of Technology,
3-7-5 Toyosu, Koto-ku, Tokyo 135-8548 Japan,
[2] International University of Health and Welfare,
2600-1 Kitakanemaru, Ohtawara City, Tochigi, Japan,
{m709102,l07104,m110013,masaomi,Ohkura}@shibaura-it.ac.jp,
ftsuchiya@iuhw.ac.jp

Abstract. The similarity of drug names in Japanese such as 'アマリール' (Amaryl) and 'アルマール' (Almarl) causes confusion over drug names and can lead to medical errors. In order to prevent such errors, methods of computing their similarity have been proposed. However, there are no studies that take account of character shape similarity quantitatively. In a previous study, we calculated the character shape similarity by template matching technique and proposed a method of measuring medicine name similarity based on it. Although we obtained a high correlation coefficient between the medicine name similarity values and subjective evaluation, we observed some character pairs that are not similar. In this study, we improved the method of computing the character shape similarity based on the characteristic points of character and compared it with advanced methods.

Keywords: Medicine name similarity, Medical safety, Character shape similarity.

1 Introduction

The similarity of drug names in Japanese causes confusion over drug names and can lead to medical errors. In Japan, accidents involving confusion over the diabetes drug, 'アマリール' (Amaryl) with 'アルマール' (Almarl), which is a drug for hypertension, have resulted in patient death. In order to prevent such accidents, the Ministry of Health, Labor and Welfare has issued notices and raised awareness among medical experts. However, errors still occur.

In order to prevent such errors, it is necessary to avoid the approval of some medical names. For this purpose, many methods have been proposed for measuring the similarity of medicine names.

Tsuchiya et al. [1] proposed similarity indices for medicine names. Based on the indices, the 'Medicine similar name search engine' [2] was developed and has been operated by The Ministry of Health, Labor and Welfare. The system measures the similarity based on the head, taking account of the existence of character pairs with a

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 628–636, 2011.
© Springer-Verlag Berlin Heidelberg 2011

similar shape and the position of the prolonged sound sign (dash) and the letter for a nasal sound in Japanese. However, although the method takes the similarity of character (letter) shapes into consideration, they assumed that the similarity of each character pair was given by hand [3].

In order to measure the similarity of character shape quantitatively and automatically, we applied the template matching method to katakana characters that compose Japanese medicine names [4]. We obtained a high correlation coefficient between the medicine name similarity values and subjective evaluation. However, we observed some character pairs in which the similarity values were low although they were similar. This is because the template matching technique does not take account of the connection between lines that compose katakana characters.

In this study, in order to solve the problem, we focused on the characteristic points of characters such as the edge points and intersection points. In the fields of character recognition, although methods based on characteristic points have been proposed, they also take account of other information such as surrounding characters, linguistic knowledge and the strokes of handwriting [5, 6]. However, since these are methods of recognizing characters rather than measuring the similarity of characters, we cannot utilize them. Therefore, it is necessary to develop a method of measuring the similarity of character shape based on the characteristic points of characters.

In this study, we propose a method of measuring the character shape similarity based on template matching technique and characteristic points. In order to evaluate the similarity index, we implement the character shape similarity to medicine name similarity and compare it with the subjective evaluation of pharmacists, which was obtained by an experiment.

2 Target Data

2.1 Medicine Names

In this study, we targeted the product names of ethical drugs that are included in the 'Standard Drug Master' provided by the Medical and Devices Agency (MEDIS-DC) in Japan [7].

The product names consist of brand, dosage form and ingredient amount. For instance, 'アマリール1mg錠' (Amaryl 1 mg tablet), 'アマリール' (Amaryl) is the brand part, '1 mg' is the ingredient amount and '錠' (tablet) denotes its dosage form. It is important to evaluate the similarity between brand parts, since pharmacists focus on brand parts when identifying a medicine. We therefore targeted only the brand part.

2.2 Character Type

The brand names of Japanese medicines are expressed in Hiragana, Katakana, Kanji characters, alphabets, numerical characters and other symbols. Among these character types, we concentrated on the Katakana characters since they are used to express many medicine names. This is because medicines are mainly named after foreign medicine names or active ingredient names.

Katakana characters are one component of the Japanese writing system and are often used to transcribe words from foreign languages. They are characterized by short, straight strokes and angular corners, for example ‘ア’(a), ‘イ’(i), ‘ウ’(u), ‘エ’(e), ‘オ’(o).

In this study, we used character images (height: 200px, width: 200px) generated by the Japanese character font, ‘MS Round Gothic’ (150 points) as source data.

3 Calculation of Character Shape Similarity

3.1 Template Matching [4]

The template matching technique is a general method that is used for image retrieval. By means of this algorithm, we can calculate the similarity value defined as the ratio of the number of the same colored pixels at the same location to the number of whole pixels. In this study, we digitalized font images, assigning zero to each white pixel and one to a black pixel, and calculated the similarity of each combination between all pairs of the target characters as shown in the following equation:

$$TM(a,b) = \frac{1}{mn}\sum_{j}^{n}\sum_{i}^{m}\delta\left(a_{ij}, b_{ij}\right), \tag{1}$$

where a and b are compared characters, m and n are the height and width of font images respectively and $\delta(x, y)$ is a function that returns 1 (if x is equal to y) or 0 (if x is not equal to y).

After the calculation, we realized that the obtained similarity values were high for all pairs: even the minimum similarity was 0.68. In order to redefine the similarity index so as to allow the minimum value to be zero and the maximum value to be one, we normalized them using the following equation:

$$sim_{tm}(a,b) = \frac{TM(a,b) - TM_{\min}}{TM_{\max} - TM_{\min}}, \tag{2}$$

where TM_{max} and TM_{min} denote the maximum and minimum value of similarity, respectively.

3.2 Characteristic Points

3.2.1 Extraction of Characteristic Points

Katakana characters have four characteristic points: edge points, folding points, branch points and intersection points. We extracted the edge points, the branch points and intersection points from the images of characters as shown in the following steps (Fig. 1). Firstly, we digitalized the font images. Secondly, we applied a thinning process to these images. Thirdly, we extracted characteristic points by pattern matching with three patterns as defined in Fig. 2. While for the other characteristic points, we extracted the folding points by hand since it is not easy to extract them.

Fig. 1. Process of characteristic point extraction

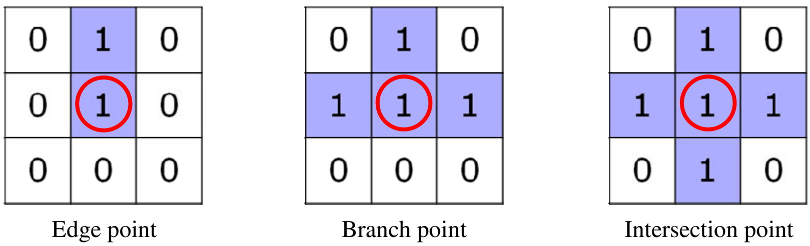


Fig. 2. Pattern of characteristic points

3.2.2 Area Division

In order to compute the value of the similarity index using characteristic points, it is necessary to find the points corresponding to the points on another image. For this purpose, we divided the font images into different areas. If the same type points exist in the same area in each image, we can regard them as correspondence points. Figure 3 shows the distribution of characteristic points extracted from all katakana characters. The figure shows that characteristic points tend to aggregate in some parts. In this study, we considered two methods of dividing the area of font images.

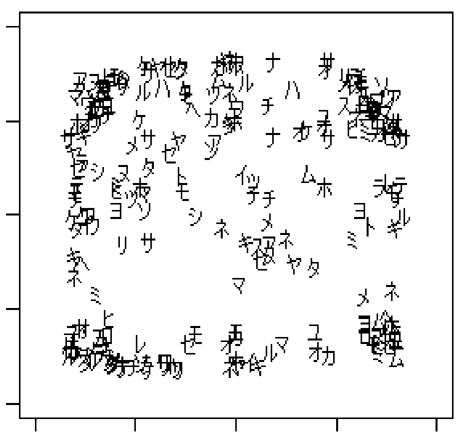


Fig. 3. Distribution of characteristic points extracted from all katakana characters

The first is a method of grid ironing the font images. The katakana characters tend to be composed of line segments extending horizontally or vertically. It is expected that dividing areas by grids takes proper account of these characters of components. However, since the most suitable grid size is not clear, we defined the size by an experiment. In addition, if there are some points near by the boundary lines of grid, they are regarded as points that belong to other areas even if they are close.

The second is a method of employing a clustering technique. The clustering technique is a method of aggregating those points into clusters that are close to each other. In this study, we utilized k-means, which is a clustering technique and aggregates the points into k clusters. By using this method, we expect that we can identify points that are close to each other to the same group. Furthermore, since the starting points and terminal points of characters tend to exist in the similar areas, it is expected to aggregate points that have the same functions. However, elongated clusters may identify separate points that should not be identified.

3.2.3 Similarity Based on Characteristic Points

If there are the same type characteristic points in the same areas, it is highly possible that they have similar characters to each other.

We calculated the absolute value of the difference between the numbers of characteristic points in an area for each type of characteristic and summed them. After computing the values in all areas, we calculated the summation of them as the distance of compared characters. Then, we divided the value by the number of all characteristic points that exist in comparing characters and subtract it from one. We defined the value as similarity based on characteristic points.

$$sim_{cp}(a,b) = 1 - \frac{\sum_D \left(|p_a^e - p_b^e| + |p_a^b - p_b^b| + |p_a^i - p_b^i| + |p_a^f - p_b^f| \right)}{\sum_D \left(p_a^e + p_b^e + p_a^b + p_b^b + p_a^i + p_b^i + p_a^f + p_b^f \right)} \quad (3)$$

where a and b are comparing characters, D are divided areas. p^e, p^b, p^i and p^f are the number of each characteristic point; edge points, branch points, intersection points and folding points.

3.3 Calculation of Character Shape Similarity

We defined the similarity index of character shape in a figure.

$$\omega_{a,b} = \alpha sim_{tm}(a,b) + (1-\alpha) sim_{cp}(a,b) \quad (4)$$

Let α be the contribution ratio between similarity values of template matching and characteristic points ($0 \le \alpha \le 1$).

4 Medicine Name Similarity

Based on the character shape similarity defined in the previous section, we computed the similarity of medicine names. In this study, we employed a method that is proposed by the advanced study [4]: Extended Letter Sequence Kernel (eLSK) and Extended Head and Tail Cosine (eHTCO).

5 Experiment

5.1 Evaluation of Character Shape Similarity

In order to evaluate our method of calculating character shape similarity, we compared it with the similarity values reported Yamade et al. [8]. In their study, they measured the similarity of katakana pairs by subjective evaluation. We use 50 katakana pairs whose similarity values are high. In order to indicate that our similarity index corresponds to a subjective view, we defined the following equation, which expresses the distance between the values of similarity, which are calculated by our method and the results obtained by Yamade's experiment.

$$D = \sum_{a,b} \left(sim_p(a,b) - sim_y(a,b) \right)^2 \tag{5}$$

where $sim_p(a, b)$ and $sim_y(a, b)$ are the similarity of the proposed method and Yamade's results between character a and b.

5.1.1 Comparison of Dividing Method

In this study, we proposed two methods of dividing the font images to compute the similarity of characters. In order to select the better method, we compared them by the evaluation index.

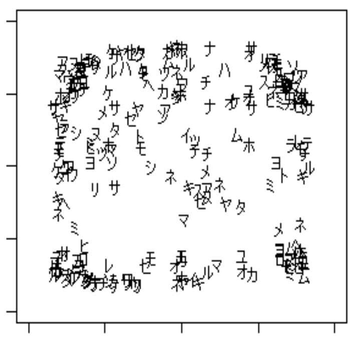


Fig. 4. Grid division (2x2)

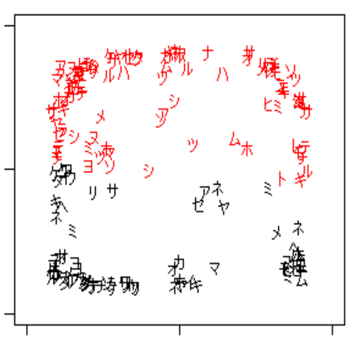


Fig. 5. Edge points (2 clusters)

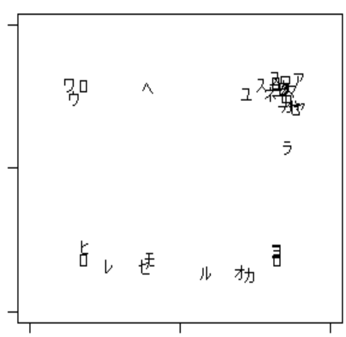


Fig. 6. Folding points (1 cluster)

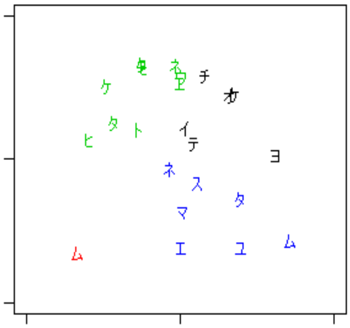


Fig. 7. Branch points (4 clusters)

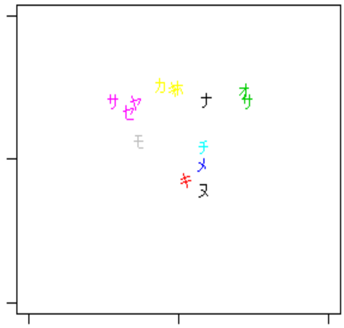


Fig. 8. Intersection points (8 clusters)

In the case of grid division, we obtained the smallest value of 1.86 when the number of horizontal division was 2 and the number of vertical division was 2 (Fig. 4). In the case of clustering division, we obtained the smallest value of 1.46 when the numbers of cluster of the edge points, the folding points, the branch points and the intersection points were 2, 1, 4 and 8 respectively (Fig. 5, 6, 7, 8). The results indicate that the latter method is better than the former method. As the reason, since most katakana characters are written from above to below or from left to right, there are edge points and folding points in the corners of font images. The superior method takes account of the features of katakana characters.

5.1.2 Comparison between Template Matching and Characteristic Point

Table 1 shows the similarity values of katakana pairs in the top 10. In the result of advanced study, we can see the dissimilar pairs whose lines match with each other such as ‘レ’ and ‘ン’. On the other hand, in the results of the proposed method, their similarity values decrease. However, there are some pairs that are not similar by the correspondence of their characteristic points such as ‘ニ’ and ‘ハ’.

Table 1. Character shape similarity in the top 10. (Yamade’s result, Template matching and Characteristic point).

A	B	Sim
シ	ツ	1.00
ソ	ン	0.94
コ	ユ	0.89
ウ	ワ	0.88
シ	ソ	0.88
チ	テ	0.88
シ	ン	0.86
ソ	ツ	0.86
ス	ヌ	0.85
ツ	ン	0.85

A	B	Sim
ク	タ	1.00
ソ	ツ	0.94
コ	ニ	0.93
エ	ニ	0.93
ク	フ	0.91
ニ	ユ	0.89
コ	ロ	0.88
シ	ツ	0.85
エ	コ	0.84
シ	ン	0.84

A	B	Sim
ア	ラ	1.00
ア	ル	1.00
コ	ワ	1.00
シ	ツ	1.00
ス	マ	1.00
ソ	リ	1.00
ニ	ハ	1.00
ヘ	レ	1.00
ラ	ル	1.00
ア	ヌ	0.88

Next, we calculated the similarity that is integrated by Equation 4. When α is 0.5, we obtained the evaluation value 0.85 by Equation 5. In addition, we can see the high correlation coefficient 0.69 between the integrated method and Yamade's result. These results indicate that it is effective to integrate the similarity indices based on the template matching and characteristic points.

5.2 Evaluation of Medicine Name Similarity

5.2.1 Method

In order to evaluate the proposed method, we compared the values computed by our method with the similarity that is evaluated by pharmacists. In the questioner, we presented the compared two medicine names on display and asked them to evaluate their similarity by values between 0 (dissimilar) and 100 (similar). In the previous study, we observed the pharmacists who answered the similarity focusing on several points such as phonological similarity. As a counter measure, we set the time (1 sec) to present the medicine names to the subjects. The number of pharmacist was 25. The names are the stem part of existing drugs. The pairs are selected so that values of similarity distribute evenly. In order to exclude the effective of length of medicine names, we selected the pairs whose medicine names are the same length. Taking account of the effect of order, we prepared 3 question patterns whose order is different from each other.

5.2.2 Results and discussion

Figure 9 shows the relation between the pharmacist evaluation and the similarity index computed by our method. These results show the correspondence between them and a height coefficient correlation of 0.87. This value is higher than the values between the pharmacist evaluation and the similarity values calculated by the method proposed in the advanced study (0.84). Furthermore, these results show the validity of the addition of similarity based on characteristic points of the existing method, which are calculated by template matching.

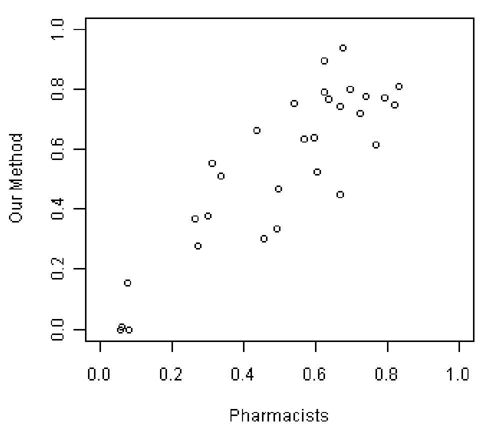


Fig. 9. Relation between pharmacist evaluation and our method

6 Conclusion

In order to ensure the medical safety of drug usage, we proposed medicine name similarity based on character shape similarity. In this study, we focused on the characteristic points of characters, which are edge points, folding points, branch points and intersection points, to redeem the method by template matching, which is proposed by the advanced study. Since it is necessary to regard points that are close to each other as corresponding points to calculate similarity using characteristic points, we proposed a method of dividing the font images into different areas. By comparing the obtained similarity values to the subjective evaluation obtained by advanced study, grid division was found to be superior to clustering division.

In order to evaluate the proposed method, we calculated the medicine name similarity based on the character shape similarity computed by the method and compared them to the similarity perceived by pharmacists. Our method was found to be superior to the advanced method, which is based on only template matching technique. Furthermore, we obtained a high correlation coefficient between our similarity and subjective evaluation of pharmacists.

In the future, it is necessary to experiment using characters that are expressed by other font types and handwriting. In addition, we should take account of the phonological similarity of characters.

References

1. Tsuchiya, F., Kawamura, N., Oh, C., Hara, A.: Standardization and similarity deliberation of Drug-names. Jpn. J. Med. Informatics 21, 59–67 (2001)
2. Japan Pharmaceutical Information Center. Medicine similar name search engine, `https://www.ruijimeisho.jp/index.aspx`
3. Ohtani, H., Takeda, M., Imada, Y., Sawada, Y.: Development of the Measures to Evaluate the Similarity of Drug Brand Names. Yakugaku Zasshi 126(5), 349–356 (2006)
4. Nabeta, K., Imai, T., Kimura, M., Ohkura, M., Tsuchiya, F.: The similarity index of medicine names to prevent confusion. In: The Pan-Pacific Conference on Ergonomics 2010 Proceedings (2010)
5. Akiyama, K., Nakagawa, M.: A Liner-Time Elastic Matching Algorithm for On-line Recognition of Handwriting Japanese Characters. The Transactions of The Institute of Electronics, Information and Communication Engineers J81-D-2(4), 651–659 (1998)
6. Nishida, S.: A Matching Algorithm of Feature Graph for Handwritten Character Recognition. IEICE Technical Report CST 108(79), 25–30 (2008)
7. The Medical Information System Development Center, `http://www.medis.or.jp/`
8. Yamade, Y., Haga, S.: Subjective evaluation of similarity of appearance of Katakana characters in drug names. Rikkyo Psychological Research 50, 79–85 (2008)

ColoriT: Color Based Image Code Application to Aid in Memory Restoration of Offline Photo Artifacts

James Park, Jonghoon Seo, Ji-Hye Choi, and Tackdon Han

Dept. of Computer Science, Yonsei University,
134, Seodaemun-Gu, Seoul 120-749, Republic of Korea
{james.park,jonghoon.seo,asellachoi,hantack}@msl.yonsei.ac.kr

Abstract. Many areas of research have applied memory aid applications to help users remember experiences or enhance learning abilities. Relatively little study however has been done on the use of color based image code as a memory aid. In fact much of what is in use by mainstream media and businesses use mobile barcodes such as 2D image code like Quick Response (QR) code and Microsoft Tag for accessing online content. Part of the freedom of using mobile code is many formats are freely based image code and have accessible API and development kits. Only few are licensed based and are limited to developing usable applications. We investigate a proprietary licensed color based image code using an application we developed called ColoriT (pronounced Color- ət) in hopes of studying its applicability to enhance memory aid when viewing photo artifacts. ColoriT is a simple photo memory tagging concept used to tag offline photos in a pervasive and natural way. By enhancing the user's ability to tag offline photos with memory artifacts we create a tool inspired by storytelling to improve memory aid and overall enjoyment of looking at photos.

Keywords: Memory aid, 2D barcode, Colorcode, Sound artifacts, Color.

1 Introduction

The use of human vision and hearing through natural features such as sound and color have been used to enhance memory cognition in various areas of research. Sound artifacts in particular can be a range of recorded voices, events, and any type of auditory inducer that triggers a memory. The use of color features can be used also to trigger a memory response usually by allowing the user to associate color with events and actions or objects. We take advantage of these natural features and incorporate them into the design of our application which helps to trigger memory responses to unfamiliar photos taken in the past. The advantage of such a system allows elderly for example to remember a particular event or person taken in a photo. Using this scenario as our motivation and inspired by storytelling and online photo tagging phenomenon we present ColoriT (pronounced Color- ət) which allows users such as elderly to remember offline photos taken in the past. Although storytelling applications [5,14] and photo tagging is not a new idea, the concept of digitally tagging offline photos with image code is relatively new. Section 3 explains the methods used in our applications as well as the initial prototyping tools used to study the natural features used in our application.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 637–642, 2011.
© Springer-Verlag Berlin Heidelberg 2011

1.1 Natural Features – Sound

Sound artifacts are a powerful tool which allows our brain to trigger memory responses. We often associate sounds with known objects in technology and objects existing in nature. Microsoft in particular has adopted universal sounds which most consumers can recognize as starting up a computer or shutting down a computer. The human mind can also distinguish auditory sounds of various objects for example different animals. We know for instance what sound a duck makes or what sound a dog produces. These bits of audio responses are stored in our memory and can collectively be recalled to associate to a particular event or object. Work focusing on using sound as a research tool to study the effects it has on memory have been used extensively. Baldis [1] uses spatial audio to enhance memory in recognizing users during a conference call meeting. A main drawback of this approach points out that visual representations do not affect the outcome in triggering correct memory responses from the test users however participants noted that having visual cues associated with audio is helpful. Dib et al [6] developed sonic souvenirs an application study to allow users to reminiscence on pictures taken in the past using a digital voice recorder. This study revealed that while applicative and enjoyable users did not know what to do with the recordings and often times were not directly tied to associated photos. This disassociation could be improved if sound and images were directly tied together. AudioMemory an application developed by Sánchez et al [16] asserts that sound can improve memory in children who are blind or have residual vision.

1.2 Natural Features – Color

Using color to enhance memory has been studied in [8,15] psychology and is actively used in [2,11,13,17] human factor and visualization studies. Spense et al [15] stresses memory can be segmented into two parts: encoding and recognition phase. Encoding is the process by which information is stored and consequently determines the performance of retrieval. We are most likely to remember an event or object when a triggered response that appears in the encoding phase is seen again in the recognition phase. Color can exploit this area and has been shown to improve memory recognition rates when used with natural scenes. Another use of color has been used in image database retrieval systems [2] a task that is tedious when thousands of images are involved. Color aid has simplified this task by allowing users to use color as an attribute in visual processing of images when searching through the database. Another recent study [13] investigates not only the use of color such as in photos but other memory aid input sources such as text, audio, and social media data. Data is sent to the user's email and allows the user to reminiscence on data received. While useful it is used mainly for online content and would be difficult to implement with paper based information such as paper-based photos.

2 Method

ColoriT is an application based on 2D barcode [12] or also referred to as image code that associates a sound artifact with an offline (paper-based) photo. Our implementation uses a well-known 2D image code known as Colorcode [3,4,9,10] however any 2D barcode could be used in the design of our system. However, like sound memory applications that help to trigger memory color associations can also help in the same way. Accounting for these features, we adopt the natural beauty of Colorcode to appeal to the human senses more than black and white based code in the design of our application. Figure 1 below describes the process of ColoriT.

Fig. 1. (From left to right) Process of using ColoriT. Family photo taken with Colorcode attached, Captured using a mobile smartphone, sound events are played back.

The first process of ColoriT extracts the data such as sound from the code captured from the mobile phone. In the second process sound artifacts linked with the code are played back to the user. In the final process, unfamiliar photos should be recognized by the triggered memory response to the recorded sound event. While any code can be used in our application colorcode's unique color scheme can be exploited to enhance the user's memory of an event or person. It is important to note that previous studies have shown color is more effective when used with natural verses non-natural scenes. One known problem existing with our implementation is the size of the code. While sound played back does not pose an issue to memory aid, many users may have difficulty in seeing the code displayed on the photo. Several solutions to this have been proposed. To counter this, larger printed code may be used however the size of code should be reasonable enough so not to pose an obvious interference with recognition of contents of photos. We assume users have adequate vision to see the color visual and we test our application with the normal 5x5 size code (approximately 7mm in width and length). Another issue is the continual effect colorcode may have to supplement sound in memory enhancement. While studies have shown more natural scenes are more efficient, we test the effects of the code in both formats (natural and non-natural). Figure 2 shows a sample of codes used in our application.

In figure 2, we include (e) as a reference to what the code might look like in black and white. It was difficult to produce scenes in black and white that were distinguishable at this size and so we leave that portion out of our experiments. Beautifying the code with natural scenes is a trivial task. The process of decoding the code uses hsv color space to compensate for the change in color intensities and we found that neighboring color cells may be overlapped up to 40% without fail to read the code using our decoder.

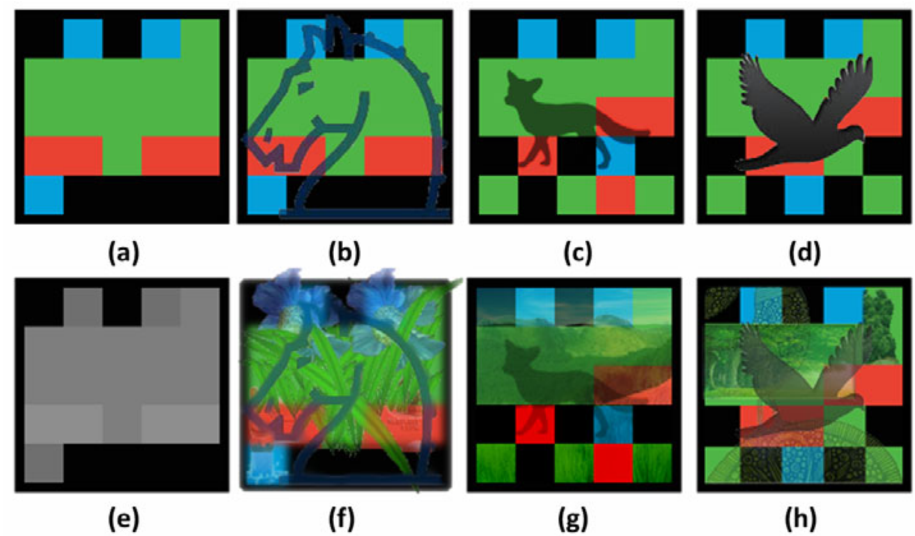


Fig. 2. (a) is an original colorcode, (b-d) are normal codes with added visual object, (f-h) are corresponding visual objects with natural scenes added. (e) shows a grey scale coded block.

3 Experiment and Results

The primary focus of our application is to test its effect on memory enhancement and the applicability of colorcode. We targeted users access to smart phones equipped with a Wi-Fi connection. Also since prior knowledge of sounds heard and known persons in photos taken is not possible to test we overcome by taking a similar approach and still maintaining the concept of our application.

Task 1: We take a random sort of photos with attached codes without natural scenes.

- This task involves using photos and associated colorcodes to test whether or not user can recognize person in the photo.
- 20 photos of five random people are shown to the user. Initially each time a photo is displayed the user looks at the photo, captures the colorcode using their mobile phone and listens to the sound.
- The process is repeated a second time and results are recorded as old or new where old means familiar and new means does not recognize.

Task 2: The same process is repeated with a new list of persons and now using code with natural scenes.

Four users participated in or experiment and revealed that while sound helped with task 1, users stated that added object within the code helped to associate a person in following line of photos. Compared with task 2 users did comparatively equal results. Out of 20 photos in task 1 average was 15 photos correct. In task 2, users did relatively the same with 16 photos correct. When asked what helped to identify with the colorcode was the object placed in the middle. Each colorcode whether with or without a natural scene contained a small silhouette of an animal. Though inconspicuous, one user stated:

"…having an animal like icon helped to distinguish between colorcodes… I also could associate colored silhouette with a person I saw before."

While it is certain natural scenes with color can enhance memory recognition, we have shown that colorcode is a great application for memory enhancement and tool for looking through paper based photos.

4 Discussion

Initial prototyping and testing of ColoriT shows that users show a better interest in the photo taken. Colors and shapes used by the code can be an element of memory aid if used properly. Actual play back of the sound happening at the time of the photo is also a natural way for users to help remember memories. Though we have shown a robust method for memory enhancing using colorcode, further studies with other media artifacts such as video and comparison with other color based image codes such as Microsoft's high capacity color barcode (HCCB) [7] is needed. The aspect of storytelling is also something to be considered. For example, when users take photographs or are creating a scrap book of photos sound clips and other media can be stored with the associating photograph at that specific moment. Codes can be created to capture the progress of one's work. This is just one avenue ColoriT can be used as ColoriT has presented itself as a natural and pervasive method in the restoration of memory and can be an effective tool to enhance memory response.

References

1. Baldis, J.J.: Effects of Spatial Audio on Memory, Comprehension, and Preference during Desktop Conferences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2001), pp. 166–173. ACM, New York (2001)
2. Chan, H.C., Wang, Y.: Human factors in color-based image retrieval: an empirical study on size estimate accuracies. Journal of Visual Communications and Image Representation 15(2), 113–131 (2004)
3. Cheong, C., Kim, D.-C., Han, T.-D.: Usability evaluation of designed image code interface for mobile computing environment. In: Jacko, J.A. (ed.) HCII 2007, Part II. LNCS, vol. 4551, pp. 241–251. Springer, Heidelberg (2007)
4. Cheong, C., Han, T.D., Kim, J.Y., Kim, T.J., Lee, K., Lee, S.Y., Itoh, A., Asada, Y., Craney, C.: Pictorial Image Code: A Color Vision-based Automatic Identification Interface for Mobile Computing Environments. In: Eighth IEEE Workshop on Mobile Computing Systems and Applications, HotMobile 2007, March 8-9, pp. 23–28 (2007)
5. Crabtree, A., Rodden, T., Mariani, J.: Collaborating around collections: informing the continued development of photoware. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, CSCW 2004, Chicago, Illinois, USA, November 06-10, pp. 396–405. ACM, New York (2004)
6. Dib, L., Petrelli, D., Whittaker, S.: Sonic souvenirs: exploring the paradoxes of recorded sound for family remembering. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW 2010, Savannah, Georgia, USA, February 06-10, pp. 391–400. ACM, New York (2010)

7. Gavin, J.: System and method for encoding high density geometric symbol set, available at United States Patent Application, 20,050,285,761 (December 2005)
8. Gegenfurtner, K.R., Rieger, J.: Sensory and cognitive contributions of color to the recognition of natural scenes. Current Biology 10, 805–808 (2000)
9. Han, T.D., Cheong, C.H., Yoon, H.M., Kim, J.Y., Jeong, S.H., Ryu, Y.S., Kang, B.S., Kim, H.K., Lee, S.W., Srini, V.P., Lee, J.H., Sohn, Y.W., Baek, Y.S., Lee, S.Y., Kang, W.S., Kim, S.W.: Implementation of New Services to Support Ubiquitous Computing for Town Life. Software Technologies for Future Embedded and Ubiquitous Systems. In: Third IEEE Workshop on Software Technologies for Future Embedded and Ubiquitous Systems (SEUS 2005), pp. 45–49 (2005)
10. Han, T.D., Cheong, C.H., Lee, N.K., Shin, E.D.: Machine readable code image and method of encoding and decoding the same, United States Patent, 7,020,327 (March 2006)
11. Henry, D., Malaiya, Y.K.: A visualization system for sliding windows protocols. In: 33rd Annual Frontiers in Education, FIE 2003, November 5-8, vol. 1, pp. T2C-1-6 (2003)
12. Kato, H., Tan, K.T.: Pervasive 2D Barcodes for Camera Phone Applications. IEEE Pervasive Computing, 76–85 (October-December 2007)
13. Peesapati, S.T., Schwanda, V., Schultz, J., Lepage, M., Jeong, S.Y., Cosley, D.: Pensieve: supporting everyday reminiscence. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI 2010), pp. 2027–2036. ACM, New York (2010)
14. Rosner, D.K., Ryokai, K.: Spyn: augmenting knitting to support storytelling and reflection. In: Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp 2008, Seoul, Korea, September 21-24, vol. 344. ACM, New York (2008)
15. Spence, I., Wong, P., Rusan, M., Rastegar, N.: How Color Enhances Visual Memory for Natural Scenes. Psychol. Sci. 17(1), 1–6 (2006)
16. Sánchez, J., Flores, H.: Memory Enhancement through audio. In: Proceedings of the 6th international ACM SIGACCESS Conference on Computer and Accessibility (Assets 2004), pp. 24–31. ACM, New York (2004)
17. Wright, P., Mosser-Wooley, D., Wooley, B.: Techniques and tools for using color in computer interface design. Crossroads 3, 3–6 (1997)

A Configurable Photo Browser Framework for Large Image Collections

Frode Eika Sandnes

Faculty of Engineering, Oslo University College, N-0130 Oslo, Norway
frodes@hio.no

Abstract. Image collections are growing at an exponential rate due to the wide availability of inexpensive digital cameras and storage. Current browsers organize photos mostly chronologically, or according to manual tags. For very large collections acquired over several years it can be difficult to locate a particular set of images – even for the owner. Although our visual memory is powerful, it is not always easy to recall all of one's images. Moreover, it can be very time consuming to find particular images in other peoples image collections. This paper presents a prototype image browser and a plug-in pattern that allows classifiers to be implemented and easily integrated with the image browser such that the user can control the characteristics of the images that are browsed and irrelevant photos are filtered out. The filters can both be content based and based on meta-information. The current version is only employs meta-information which means that large image collections can be indexed efficiently.

Keywords: Photo browsing, content based image retrieval, meta-information.

1 Introduction

High quality digital cameras are inexpensive and within the reach of most people. Storage capacity has grown exponentially and prices have dropped. Moreover, most mobile devices are equipped with powerful cameras. People therefore take more photographs than ever before, and image collections are therefore growing at an exponential rate.

Most camera owners have a have a gut feeling about what content they have in their personal collection and approximately where they can find particular images. However, for shared image repositories where multiple people contribute images and use the photographs it can be a huge challenge to find exactly the images one is looking for. The literature focuses on three approaches to handling large image collections, namely manual image tagging [1, 2], content based image retrieval [3-5] and image browsing [6-8]. Manual image tagging is the strategy whereupon an image or a collection is manually tagged with keywords or text. Tagged images can be retrieved using text-based queries. Many camera owners also organize their amateur images into folders where each folder is given a descriptive name. Although crude, the folder names can be useful when searching for images. Unfortunately, manual image tagging is very time consuming especially with large collections. Labeling of

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 643–652, 2011.
© Springer-Verlag Berlin Heidelberg 2011

folders is more manageable than tagging individual photos. Another problem with manual tagging is that it is a subjective process dependent on how the observer perceives the image content, and two different people may tag images differently. In practice, however, people do not voluntary tag their images and several researchers have experimented with games to encourage the manual tagging of images [1, 2]. Several automatic tagging strategies have also been proposed such as the strategy used by the well known service Google image search, whereupon images on web pages are associated with the text on the pages in which they appear. Several strategies are also based on the automatic analysis of image contents.

Content based image retrieval is an active research topic where various approaches for automatically identifying the contents of images have been proposed [3, 9]. Content based image retrieval is a very complex process and the successful examples have been limited to special domains such as medical [5], landmark recognition [9] and general shapes [4] to mention at few.

Image browsing is yet another area that has received some attention, especially in the field of human computer interaction [10-12]. A browser allows the user access to all the images. As the human visual system is good at scanning, browsing appears to be a preferred method of accessing images as the user is in control. However, with large collections even browsing can be a challenge. Most commercial browsers are very simple in their approach, as they organize images chronologically and typically exploit folder names. Browsing applications such as Picasa has also employed techniques from image processing, in particular face recognition, to assist in the image browsing process. The merging of ideas from browsing and image processing is an emerging trend [6-8] and this is also the basis for the approach taken in this paper.

2 Ideation

The background for this work was the author's previous research on techniques for extracting geographical information from image collections [13-16] where the geographical information is extracted from the image meta data stored in the image EXIF headers [10, 17, 18]. This has for instance been applied to automatic tagging of images and providing summaries of collections in the form of image browsers for the blind [19].

The initial work started with a video sketch of a touch based image browsing application for children shown in Fig. 1. In this browser the child will construct a scene or scenario representing the images the child wants to see. The child is represented by a character that can be placed indoors or outdoors, representing indoor images or outdoor images, what time of day, that is, morning, noon, afternoon, evening and night, when the scene is inverted to represent the darkness of night. The weather conditions are indicated by flicking the sun where the user cycles through clear skies, partially cloudy days and cloudy days. Moreover, the season can be indicated by stretching or shrinking the sun. If the sun is shrunk it becomes winter and the house is covered in snow. Also the geographical location of the images is represented by the characters clothing, here represented by a Viking from northern Europe, a cowboy from America, a Chinese outfit, etc. The clothing is changed by flicking the character. The thumbnail images display would automatically change when scenario is changed. This interface is not dependent on text or technical characteristics of images, hence, making the interface suitable for children. The video sketch was created by first making the animated cartoon illustrating the various screens, then played back on an ordinary computer with the hand gestures enacted on top and the resulting scene filmed with a video camera.

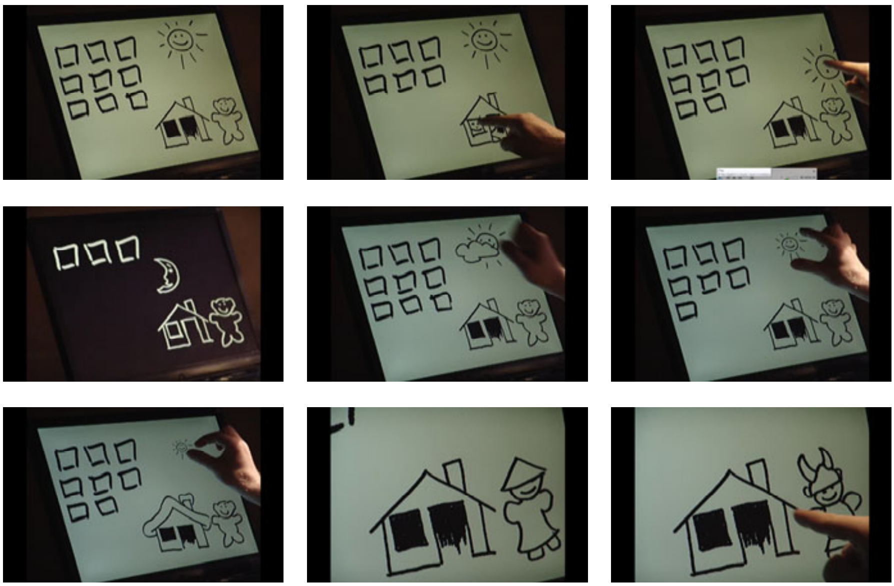


Fig. 1. A video sketch of a photo browser for children. a) the start screen, b) move the character indoors, c) changing the time of day, d) at night, e) chaning the weather conditions, f) shrinkning the sun, e) shrinkning the sun into winter, f) chaning the geographical location into South East Asia, g) chaning the geographical location into Northern Europe.

3 Image Browsing Framework

Fig. 2 shows an overview of the image browsing framework. The user is exposed to the graphical user interface that can be tailor made for a particular domain such as the children's image browser shown in Fig. 1. The GUI is built on top of the image browsing framework which maintains a database of all the images in a given collection. The image browsing database helps speed up browsing such that all the attributes of interest connected to a given image is easily retrievable and the interface can respond to user queries in real-time. Therefore, each time an image is browsed the system does not need to analyze the image. Images are only analyzed once when loaded into the system the first time.

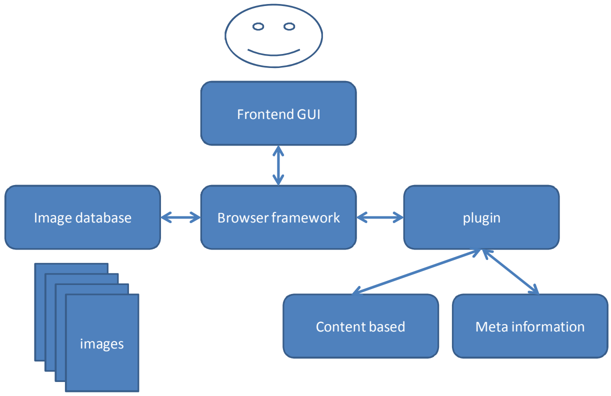


Fig. 2. Framework architecture

Fig. 3. The photo browser without any criteria selected

3.1 Content-Based versus Meta-information Based

The plug-ins are written for the framework, and a plug-in can be content based or based on meta-information. Image content based plug-ins typically analyses the image contents as a basis for providing a classification. Content based analysis can be everything from simple analysis of average pixel intensity and mean pixel hue to more complex image analysis such as object recognition. Content analysis is time-consuming, especially with high resolution images as images as analyzed at pixel level. The complexity of the image analysis algorithm will significantly affect the processing time, and the more sophisticated object recognition algorithm for instance, can have a high time-complexity. Moreover, the processing time is proportional to the size of the collections and the processing of a very large collection can therefore be computationally demanding.

Image meta-information can provide a vast range of valuable information that can be analyzed. Modern digital cameras store the state of the camera in so called EXIF headers [10, 17, 18], for example the shutter speed, aperture, film speed (ISO), the time and date the image was taken, lens focal length, etc. Some of these settings such as shutter speed, aperture and film speed can be used to compute more abstract quantities such as the exposure value [20]. The exposure value gives an absolute measurement of the light intensity at the scene and such information cannot be extracted fully from the content of images as the image measurements are relative. Moreover, even more abstract information such as geographical location [16] and descriptive information relating to local time zones [19] can be extracted based on these quantities. One strength of meta-information is that it can be exploited without having to inspect the image contents and is thus not related to the image size. Meta-information analysis is typically much less computationally intensive than content based analysis.

The framework supports a mixture of content and meta-information based image classifiers.

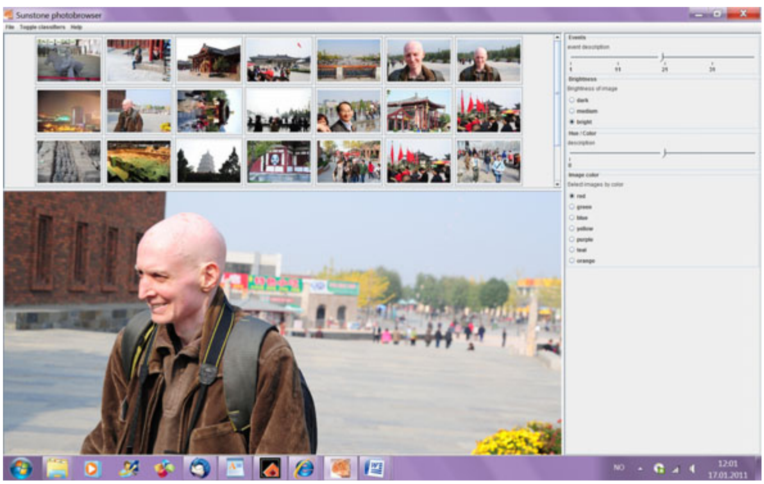


Fig. 4. The photo browser with the settings "bright" and "red"

3.2 Image versus Collections

The framework supports two further types of plug-ins namely individual image classifiers and collection classifiers. An individual image classifier is responsible for analyzing a single image and extracting a particular feature. For example, this could be the overall image intensity, mean hue, the month the image was taken, etc. However, features such as brightness level are characteristics that are only meaningful to connect to single images and not necessarily a collection of images. On the other hand image collection classifiers finds characteristics common to a set of images. For example, images can be grouped according to their temporal spread, that is, images taken within the same few days could be part of the same holiday, while images separated by more than a few days could belong to totally unrelated events. Moreover, images taken within one related event are typically taken in a limited geographical region, time-zone, etc. A collection classifier will thus typically classify groups of images.

3.3 Extendibility

The plug-ins are designed to take an image or image collection as input, that is, they have access to all the image characteristics, contents, and meta-information, as well as the output of other plug-ins, and then produce a single feature as output. A separate plug-in is thus needed for each feature that should be represented in the interface. Therefore, there is a precedence relationship between the plug-ins dictating their dependencies.

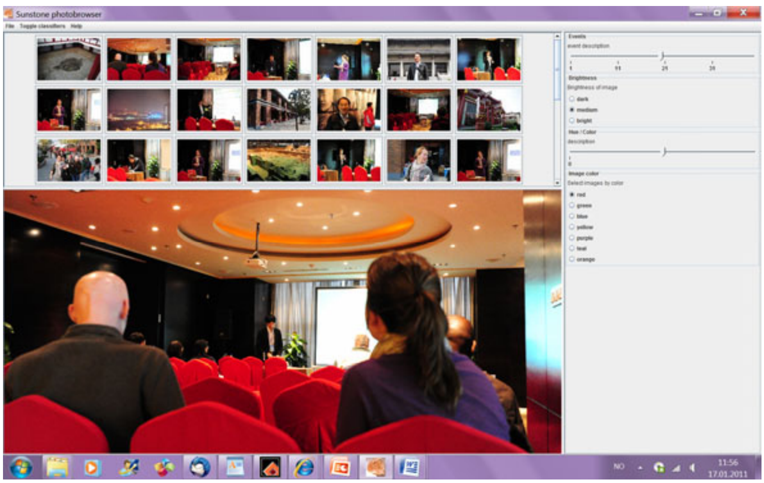


Fig. 5. The photo browser with settings "red" and "medium"

The output of each plug-in is represented as a visual control in the user interface. For the default interface, implemented in the current framework this is either a category represented as a pull down menu, choices represented as radio buttons, or a slider representing a quantity. It is also possible to represent support plug-ins whose values are not displayed in the interface, but provided as input to other classifiers.

Each plug-in is run on each image once to expedite short response times. Once a plug-in has categorized an image the corresponding value is stored in the database. Plug-ins and images can be added dynamically, and once a new image is added, all the plug-ins are applied to that image. Likewise, if a new plug-in is added, the plug-in is applied to all the images.

The framework is illustrated in Figs. 3-5, 7 and 8. The criteria for the images are selected in the rightmost pane, the corresponding selection of images are shown in the top left pane and the selected image at any given time is shown in the bottom left pane.

```
public ClassifierDayNight()
       {
       super( false,
              new Component("Day or night",
                            "Show photos of day or night",
                            new String[]{"Day",
                                          "Night" }. true ));
       }
  Public   Object   processElement(HashMap<String,   Object>
input)
       {
       double EV = 0;
              {
              Object                     tmp                    =
input.get(exposureValueClassifier);
              if(tmp instanceof Double)
                     {
                     EV = (Double) tmp;
                     }
              else
                     {
                     throw new NullPointerException(
                                          "Need Exp. Value”);
                     }
              }
       return (EV >= 11);
       }
```

Fig. 6. Extract of a classification filter implementation

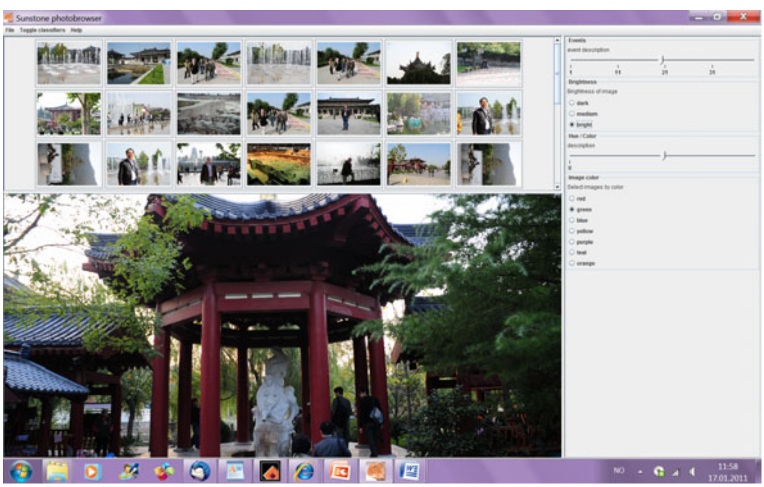


Fig. 7. The photo browser with settings “green” and “bright”

The framework is written in Java. Plug-ins are also written in Java and the constructors are used to signal whether it is a collection classifier or single image classifier as well as its appearance in the GUI, with explanatory text. Extracts from an example classifier plug-in is shown in Fig. 6.

The various plug-ins can be added dynamically by adding their respective class-files to the application. However, they require a restart as the database is checked against the available images and plug-ins only during startup. During run-time the user can activate or de-activate various plug-ins via the user interface. This allows the user to display only the settings that are relevant in a given application and unnecessary clutter is avoided.

4 Demonstration

The early prototype has not yet been evaluated on a panel of users, as the set of implemented classifiers are limited and the front end is yet crude. Figs. 3-5, 7 and 8 illustrate the prototype photo browser on a collection of 480 images taken at Xi'an China during the UIC/ATC'2010 conference in October 2010. Fig. 3 shows the browser when started. Initially all the images are displayed as no criteria are selected. In Fig. 4 the browser is set to show "red" images that are "bright". Red images are images with some amount of red in them. Clearly, there is only a fraction of the images that satisfies these criteria which can be seen from the scrollbar. Next, in Fig. 5 the setting "red" with "medium" intensity is selected. Again, the set of images is much smaller and many of the images are indoors. Fig. 7 and 8 shows the images that are displayed when selecting "green" and "blue" images, respectively – both in the category "bright".

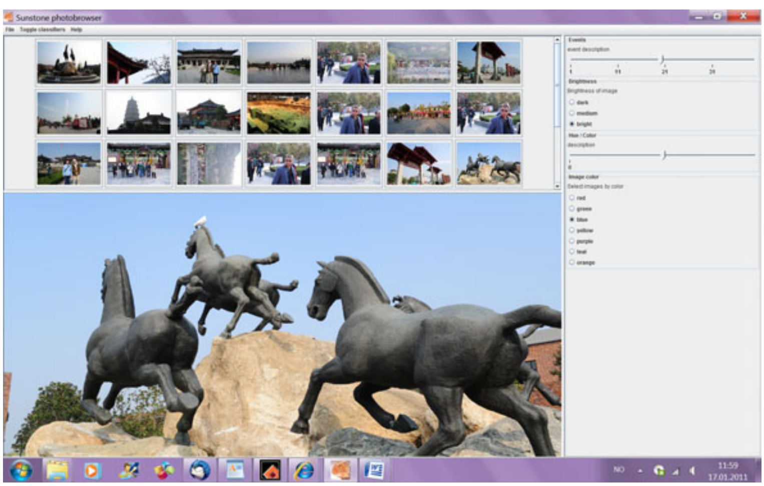


Fig. 8. Image browser with selections "bright", "blue"

5 Conclusions

This paper has presented a customizable image browser with an architecture that allows plug-ins to be designed for specific purposes based on image contents or image meta-information, or both. This strategy holds potential for browsing large image collections and especially image collections where the viewer is not the photographer. The strategy bridges techniques and the benefits from traditional image browsing with the recent advances in image processing and analysis. Future work involves testing the framework with novel image sorting categories on target groups such as children.

Acknowledgements. The author is grateful for the hard work of Martin Haukeli, Jostein Haukeli and Marit Olsen who implemented the Sunstone browser and contributed to the design of the plug-in pattern.

References

1. Diakopoulos, N., Chiu, P.: Photoplay: A collocated collaborative photo tagging game on a horizontal display. Presented at UIST 2007 (2007)
2. Golder, S., Huberman, B.A.: The structure of collaborative tagging systems. Journal of Information Sciences 32(2), 198–208 (2006)
3. Carboni, D., Sanna, S., Zanarini, P.: GeoPix: image retrieval on the geo web, from camera click to mouse click. Presented at Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services (2006)
4. Huang, Y.-P., Chang, T.-W., Sandnes, F.E.: Efficient Shape-Based Image Retrieval Based on Grey Relational Analysis and Association Rules. International Journal of Pattern Recognition and Artificial Intelligence 22(4), 711–732 (2008)
5. Quellec, G., Lamard, M., Cazuguel, G., Cochener, B., Roux, C.: Wavelet optimization for content-based image retrieval in medical databases. Medical Image Analysis 14(2), 227–241 (2010)
6. Bruneau, P., Picarougne, F., Gelgon, M.: Interactive unsupervised classification and visualization for browsing an image collection. Pattern Recognition Letters 43(2), 485–493 (2010)
7. Chu, W.-T., Lin, C.-H.: Consumer photo management and browsing facilitated by near-duplicate detection with feature filtering. Journal of Visual Communication and Image Representation 21(3), 256–268 (2010)
8. Jin, J.S., Kurniawati, R., Xu, G., Bai, X.: Using Browsing to Improve Content-Based Image Retrieval. Journal of Visual Communication and Image Representation 12(2), 123–135 (2001)
9. Zheng, Y.-T., Ming, Z., Yang, S., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., Neven, H.: Tour the world: Building a web-scale landmark recognition engine. Presented at IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009) (2009)
10. Jang, C.-J., Lee, J.-Y., Lee, J.-W., Cho, H.-G.: Smart Management System for Digital Photographs using Temporal and Spatial Features with EXIF metadata. Presented at 2nd International Conference on Digital Information Management (2007)
11. Schaefer, G.: A next generation browsing environment for large image repositories. Multimedia Tools and Applications 47(1), 105–120 (2010)

12. Porta, M.: Browsing large collections of images through unconventional visualization techniques. Presented at AVI 2006 (2006)
13. Sandnes, F.E.: Sorting holiday photos without a GPS: What can we expect from contents-based geo-spatial image tagging? In: Muneesawang, P., Wu, F., Kumazawa, I., Roeksabutr, A., Liao, M., Tang, X. (eds.) PCM 2009. LNCS, vol. 5879, pp. 256–267. Springer, Heidelberg (2009)
14. Sandnes, F.E.: Determining the Geographical Location of Image Scenes based on Object Shadow Lengths. Journal of Signal Processing Systems for Signal, Image, and Video Technology (2010)
15. Sandnes, F.E.: Unsupervised and Fast Continent Classification of Digital Image Collections using Time. Presented at ICSSE 2010 (2010)
16. Sandnes, F.E.: Where was that photo taken? Deriving geographical information from image collections based on temporal exposure attributes. Multimedia Systems 16(4-5), 309–318 (2010)
17. Alvarez, P.: Using Extended File Information (EXIF) File headers in Digital Evidence Analysis. International Journal of Digital Evidence 2(3) (2004)
18. Romero, N.L., Chornet, V.V.G.C.G., Cobos, J.S., Carot, A.A.S.C., Centellas, F.C., Mendez, M.C.: Recovery of descriptive information in images from digital libraries by means of EXIF metadata. Library Hi Tech. 26(2), 302–315 (2008)
19. Sandnes, F.E.: A Non-Visual Photo Collection Browser based on Automatically Generated Text Descriptions. Presented at The Third International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2010 (2010)
20. Ray, S.F.: Camera Exposure Determination. In: Jacobson, R.E., Ray, S.F., Atteridge, G.G., Axford, N.R. (eds.) The Manual of Photography: Photographic and Digital Imaging, Focal Press (2000)

Imaged Based Codes Performance Comparison for Mobile Environments

Jonghoon Seo, Ji Hye Choi, and Tack-don Han

134 Sinchon-dong, Seodaemun-gu, Seoul, Korea
{jonghoon.seo,asellachoi,hantack}@msl.yonsei.ac.kr

Abstract. By spreading of smart phones, mobile barcodes are used widely. However, there are so many 2D barcodes to be available. So, it is important to compare those mobile barcodes. In this paper we performed decoding performance comparison between popular mobile barcodes. ColorCode is using color information to get information. So it shows most improved performance in distance and size. Also, it can provide magnifying decoding mode, and it shows more enhancing result.

Keywords: 2D Barcode, Image based Code.

1 Introduction

The spread of smartphones is expanding as the digital information access becomes available in the mobile environment. Especially, because of uncomfortable entering devices, using the Tag Interface is simple and faster to access the information that person wants. Among those tag interfaces, the 2D barcode is being used a lot because of the advantage that can easily access to it.[1] The 2D barcode is developed to overcome the capacity limitation. Hence, it was researched on design and recognition of high-capacity code in the beginning, but other area of study such as the feature for the user point of view became an important study area. Unlike previous 2D barcode market, users access to information by recognizing the code that is located at the long distance in the mobile code recognition. Therefore, the performance comparison about recognition distance for each code is needed. Also, the barcodes obstructs the existing contents. To minimize the bothering, the smaller size of the code the better. For this, it is necessary to measure the smallest size of each code by comparing the minimum size of code that can be recognized.

In this paper, we performed the capacity comparison which is required in mobile code recognition. In capacity comparison, we compared the capacity between QR Code Data Matrix, and ColorCode that are mostly used in Korea mobile code market. The capacity comparison was performed on the comparison of the minimum size of code that can be recognized and the distance that code can be recognized. Through these experiment data, we can use them as the guideline to utilize the mobile code in real life.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 653–659, 2011.
© Springer-Verlag Berlin Heidelberg 2011

In addition, in the case of ColorCode, the magnifying recognition mode is supported. It is a mode that supports long distance recognition by digital zoom the video manually. Since ColorCode uses the color and recognizes the ID, the size of one cell is bigger than B/W code. With this, the information of cell can be protected even though in the case of long distance video was digital zoomed. On the other hand, B/W codes uses B/W pattern to perform the recognition. Therefore, it was hard to recognize in despite of magnified because the cell was blurred.

In this paper, we will analyze the performance of mobile code through these experiments and provide guide line of recognition to business people and mobile code users. We will explain the 2D barcode technology and mobile code technology on second chapter, the performance result and experiment design on third chapter, talks about the experiment about magnifying recognition mode of ColorCode in chapter 4, and conclude on chapter 5.

2 Mobile Barcode

The previously developed codes can be divided into 1D barcode and 2D image code. 2D image code can be divided into B/W code which only uses black & white colors and the ColorCode which make the record density high by providing the colors to code.

In the case of 1D barcode, with the fast recognition speed and the correct recognition with free direction, it is being used in many industries. However, because of some information expression limitation and low density of recording information, there are some disadvantages that the size of code gets bigger to express lots of information[2, 3].

For the 2D image code, it is developed for improving the amount of storage problem of 1D barcode[4]. The various codes have been developed, but there was a problem that they had to use the expensive reader to decode. The examples for these 2D image codes are QR code[5] from Denso which can be read fast by finder pattern, the PDF 147[6] from Symbol Technologies which has a high rate of restoration when it got damaged, the Data Matrix from International Data Matrix[7] that is available to microminiaturize and high density of recording, the Maxi Code[8] from UPS that was developed for fast classification of parcel service cargo, the CyberCode[9] from Sony that was developed for augment reality, Ultracode from Zebra, ColorCode[10] from ColorZip Media, and Microsoft's HCCB[11]. These codes have an advantage on the amount of recording, types of data, and improved high density of record. From among these, ColorCode , Ultracode, and HCCB are using colors so that it can express more information on the same area.

Especially, the codes that can be recognized with normal camera not with the dedicated reader are spreading it as the mobile code. These mobile codes are being used as the connection of webpage for the product, advertisement, and information. Therefore, it is more likely to focusing on the usability on mobile instead of the capacity of storage which was the main focus of 2D barcode technology in the beginning.

We classify these codes with following features.

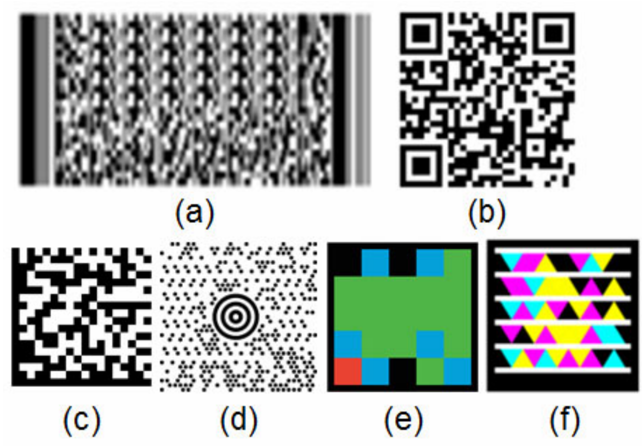


Fig. 1. Various 2D Barcodes. (a) PDF417, (b) QR Code, (c) Data Matrix, (d) MaxiCode, (e) ColorCode, (f) HCCB.

Table 1. Classification of Barcodes

	1st Generation	2nd Generation	3rd Generation
Code	1D Barcode	2D High Capacity Barcode	2D Mobile Code
Feature	Accuracy	Capacity, Speed	Color, Mobility
Example	EAN/UPC, Codabar	QR Code, PDF417, Data Matrix, CyberCode, etc	QR Code, Data Matrix, HCCB, ColorCode, etc

3 Experiments and Analysis

In this paper, we performed a comparison using three mobile codes - QR Code, Data Matrix, and ColorCode. Our experiment results show the recognition performance. For this experiment, we used iPhone as a detecting device, AT&TScanner application for Data Matrix and QR Code, and the Colorzip application for ColorCode. It is the experiment about the mobile code so all three code were encoded to have URL information. We put up the codes on the LCD monitor to measure the changes of size easily.

3.1 Minimum Size Experiments

Minimum Size from Fixed Distance. 2D barcode has many advantages, but there is a problem that interrupts the user and covers the existing media content because it is printed in visually[12]. To provide the same function, it is preferred to print the code in minimum sizes to not to cover the existed content. Therefore, we performed an experiment to examine the minimum size of the code that be recognized.

For the first experiment, we measured the minimum sizes that can be recognized in the fixed distance (50 cm). This experiment is important because it can measure the recognition performance for code itself. We calculate the average with 10 repeated experiment performed.

Table 2. Minimum size of decodable codes in a fixed distance

Fixed Distance (50cm)	Data Matrix	QR Code	ColorCode
Minimum Size (mm)	20.4 x 20.4	31.0 x 31.0	14.8 x 14.8

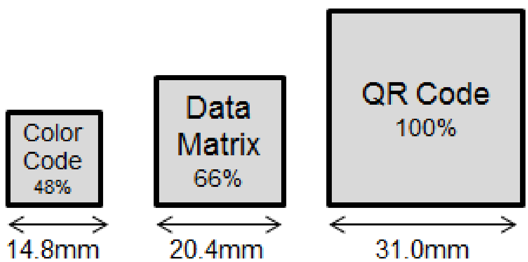


Fig. 2. The comparison of relative size (When we set QRCode as 1)

The ColorCode was able to recognize with the half size of QRCode. This can be analyzed that ColorCode increased the data integration by using the colors. The reason why Data Matrix can be recognized with the small size than QR Code is because the area of the bit amount of Finder Pattern area.

Minimum Size from free distance. The next experiment considers the mobile code recognition environment. When the user decodes the code in the real mobile life, users hold their mobile phone with hand and tries in the most appropriate angle and the distance that can be recognized easily. This experiment assumes that experimenter is in this situation and tries to decode the code with the comfortable location and the angle and measures the smallest size of code that can be recognized. As mentioned above, we experimented 10 times and calculated the average.

Table 3. Minimum size of decodable codes in a free distance

Free Distance	Data Matrix	QR Code	ColorCode
Minimum Size (mm)	7.8 x 7.8	10.4 x 10.4	2.6 x 2.6

In the case of ColorCode the minimum size of the code that can be recognized was 2mm x 2mm. By using the color, it showed as the increase of the cell's integration and very little area was needed compared to other codes. This means that it can have the information with the 1/4 size of QR Code. The smaller area was reported in the free distance experiment than the fixed distance experiment and the ColorCode had a

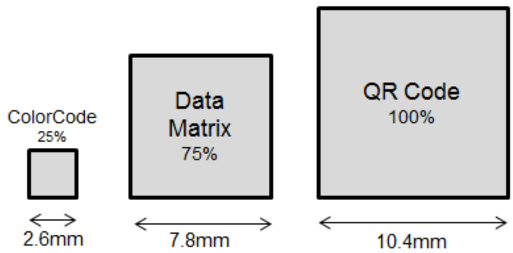


Fig. 3. The comparison of relative size (When we set QRCode as 1)

bigger gap than any other codes. It is analyzed as the cause of increasing recognition rate because of the good preservation of colors in the everyday environment that users make to photograph well.

3.2 Recognition Area Experiments

The next experiment is using the fixed size of the code and measured the area that can be recognized. For the mobile code, there is an advantage in usability that can be recognized in the long distance. Additionally, it is necessary for users to not only just recognizing at the front but also the recognition in the broad angle to make the code available to recognize on the side. We did the experimentation with the various angles and the code which has a size of 100mm x 100mm.

Table 4. Recognizing area of each codes

Recognizing Area	Data Matrix	QR Code	ColorCode
Maximum Distance(m)	~1.4	~1.3	~2.0
Maximum Angle(‘)	~56	~61	~32

As you can see in this table, the code that has the longest recognition range was the ColorCode. However, for the recognition angle, QR Code was most tolerant. This can be analyzed that in case of ColorCode, the colors has been changed because of light that was chaged by the angle.

4 ColorCode Magnify Decoding Mode

In ColorCode, the magnifying recognition mode is being supported with digital Zoom function. This mode performs on decoder when it selects “Use Long Distance” like figure 4. This mode uses the method that performs the recognition by moving or Digital Zooming after photographing like figure 4-b.

Since B/W codes recognizes B/W’s pattern of cell, the cell pattern becomes blurred when the size gets smaller or the distance increased.

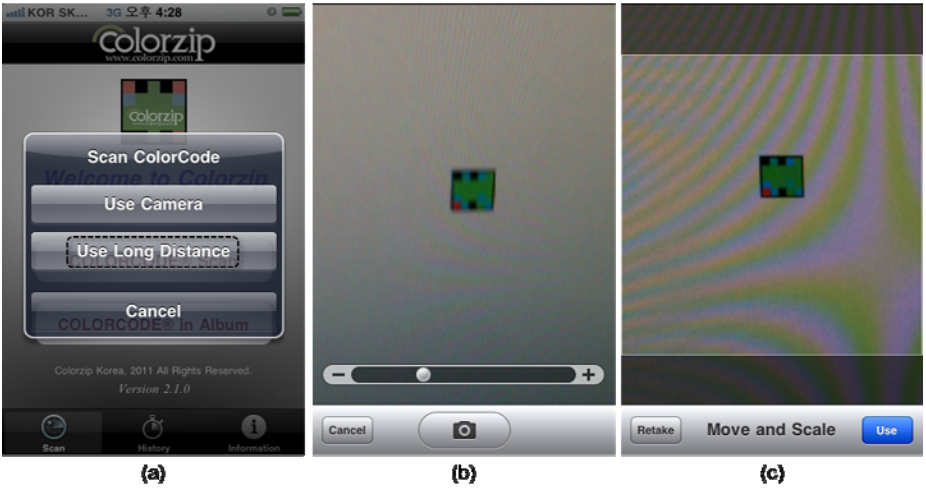


Fig. 4. ColorCode Magnify Decoding Mode

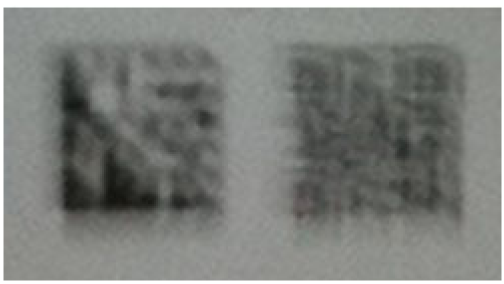


Fig. 5. Blurred Image of B/W Codes - Data Matrix(Left) and QRCode(Right)

On the other hand, ColorCode uses colors so it keeps the information of cell in the long distance since the color increases the integration of each cell. The recognition method that uses this feature is called “magnifying recognition mode.” When users use magnifying recognition method, the users need to input but the advantage that the recognition dramatically exists.

The result of magnifying recognition mode is notifying on table 5.

Table 5. ColorCode’s magnifying decoding performance

Recognizing Area	ColorCode (100mm x 100mm size)
Maximum Distance(m)	~4.70
Maximum Angle(‘)	~57

5 Conclusion

In this paper, we performed a comparison experiment between Data Matrix, QR code, and ColorCode that are mostly used for the mobile code. The main purpose for these mobile codes is to access web sites through mobile phones. The mobile codes that are used currently provide these kinds of services, the usability in the recognition process become important issues instead of capacity issue. This paper measures both the minimum size and the maximum recognition distance of code for the usability. ColorCode uses color cells so there were advantages on the aspect of the distance and size. Additionally there was considerable advantage on the distance than any other codes since ColorCode provides the magnifying recognition mode.

References

1. Kim, D.-C., Seo, J.-H., Chung, C., Han, T.-D.: Tag Interface for Pervasive Computing. In: Signal Processing and Multimedia (SIGMAP 2006), Setubal, Portugal (2006)
2. Pavlidis, T., Swartz, J., Wang, Y.P.: Fundamentals of Bar Code Information Theory. IEEE Computer 23(4), 74–86 (1990)
3. General EAN.UCC Specification v.6.0. (February 2007), http://www.ean.se/GSV6.0/HTML_Files/Document_Library/Bar_Code/00103/index.html
4. Pavlidis, T., Swartz, J., Wang, Y.P.: Information Encoding with Two-Dimensional Bar Codes. IEEE Computer 25(6), 18–28 (1992)
5. International Organization for Standardization: Information Technology – Automatic Identification and Data Capture Techniques – Bar Code Symbology – QR Code. ISO/IEC 18004 (2000)
6. International Organization for Standardization: Information technology - Automatic identification and data capture techniques - Bar code symbology - PDF-417. ISO/IEC 15438 (2001)
7. International Organization for Standardization: Information technology - Automatic identification and data capture techniques - Bar code symbology - Data Matrix, ISO/IEC 16022 (2000)
8. International Organization for Standardization: Information technology - Automatic identification and data capture techniques - Bar code symbology - MaxiCode, ISO/IEC 16023 (2000)
9. Rekimoto, J., Ayatsuka, Y.: CyberCode: designing augmented reality environments with visual tags. In: Proceedings of the 2000 ACM Conference on Designing Augmented Reality Environments (DARE 2000), pp. 1–10. Association for Computing Machinery, Helsingør (2000), doi:10.1145/354666.354667 (retrieved 2008-02-04)
10. ColorCode. ColorZip Media Inc. (February 2007), http://www.colozip.com
11. Microsoft Research. High Capacity Color Barcode Technology (retrieved December 12, 2009)
12. Cheong, C., Kim, D.-C., Han, T.-D.: Usability Evaluation of Designed Image Code Interface for Mobile Computing Environment. In: Jacko, J.A. (ed.) HCII 2007. LNCS, vol. 4551, pp. 241–251. Springer, Heidelberg (2007)

A Global Optimal Algorithm for Camera Calibration with One-Dimensional Objects

Liang Wang[1], FuQing Duan[2], and Chao Liang[1]

[1] College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, P.R. China
[2] College of Information Science and Technology, Beijing Normal University, Beijing 100875, P.R. China
wangliang@bjut.edu.cn

Abstract. The emergent one-dimensional (1D) calibration is very suitable for multi-camera calibration. However its accuracy is not satisfactory. Conventional optimal algorithms, such as bundle adjustment, do not perform well for the non-convex optimization of 1D calibration. In this paper, a practical optimal algorithm for camera calibration with 1D objects using branch and bound framework is presented. To obtain the optimal solution which can provide ε-optimality, tight convex relaxations of the objective functions are constructed and minimized in a branch and bound optimization framework. Experiments prove the validity of the proposed method.

1 Introduction

Camera calibration, which is to determine camera parameters, is a necessary step to extract three-dimensional(3D) metric information from two-dimensional (2D) images. According to the dimension of the calibration objects, the camera calibration techniques can be roughly classified into four categories: 3D object based calibration[1,2], 2D plane based calibration[3,4], one-dimensional (1D) object based calibration[5,6,7,8,9] and zero-dimensional (0D) approach or self-calibration [10,11,12]. Much work has been done expect for 1D calibration. Camera calibration using 1D object was proposed by Zhang in [5]. Hammarstedt et.al. analyze the critical configuration of 1D calibration and provide simplified closed-form solutions in Zhang's setup [6]. Wu et.al. prove that the rotating 1D object used in [5] is essentially equivalent to a familiar 2D planar object, and such equivalence still holds when the 1D object undergoes a planar motion rather than the rotation around a fixed point [7]. Wang et.al. present an 1D calibration algorithm for multi-camera in which the 1D object can undergo general motions without special constraints [8]. The emergent 1D calibration has the following advantages: (1). The 1D calibration object is easy to construct. In practice, it can be constructed by marking three points on a stick. (2). The 1D calibration is very suitable for multi-camera calibration, in which the whole 1D calibration object can be simultaneously observed by all cameras, whereas 3D or 2D calibration objects hard satisfy.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 660–669, 2011.
© Springer-Verlag Berlin Heidelberg 2011

The deficiency of 1D calibration is that its accuracy is inferior to that of 3D and 2D calibration. One reason is that the 1D calibration uses the co-linearity and distance of markers on the 1D segment to compute the intrinsic matrix, which means only one dimension of 3D metric information is used. Some researchers propose some methods and resort to nonlinear optimization to refine 1D calibration results [9]. However these methods are not guaranteed to perform well since the 1D calibration optimization is a non-convex nonlinear optimization problem.

In this paper, we present a practical algorithm for 1D calibration of multi-camera system with the theoretical guarantees of global optimality. Given images captured by a camera, firstly image points of marks on 1D calibration objects are extracted with image processing techniques and initial calibration results are obtained with original 1D calibration algorithm. Normalize the initial calibration results and use them to give bounds on variables. Secondly construct tight convex relaxations of the objective functions and minimize the objective functions in a branch and bound optimization framework. Then normalized camera parameters can be obtained by the decomposition of the dual image of the absolute conic. Finally, the camera parameters can be obtained by de-normalization.

The paper is organized as follows. Some preliminaries are introduced in Section 2. In Section 3 the proposed global optimal calibration algorithm for a camera is presented. Then calibration experiments are reported in Section 4. Section 5 are some concluding remarks.

2 Preliminary

2.1 Camera Model

In this paper, a 3D point is denoted by $\mathbf{M} = [X, Y, Z]^T$, and a 2D image point by $\mathbf{m} = [u, v]^T$. The corresponding homogeneous vector is denoted respectively as $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$, $\tilde{\mathbf{m}} = [u, v, 1]^T$. With the pinhole camera model, the relationship between a 2D image point and $\tilde{\mathbf{m}}$ its corresponding 3D point$\tilde{\mathbf{M}}$ is

$$d\tilde{\mathbf{m}} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\tilde{\mathbf{M}}, \quad \mathbf{K} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where d is a scale factor (the projection depth of 3D point $\tilde{\mathbf{M}}$), $\mathbf{K}$ is called the camera intrinsic matrix, α, β denotes the scale factors along the u and v image axes respectively, γ the skew, and $[u_0, v_0]$ the principal point, $[\mathbf{R}|\mathbf{t}]$ called the extrinsic matrix, $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ translation vector which relates the world coordinate system to the camera coordinate system.

2.2 1D Calibration Object

In this paper, the minimal configuration of 1D calibration object which only consists of three collinear points is considered. Assume three points of 1D calibration object are $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, where $||\mathbf{A} - \mathbf{C}|| = L_1$, $||\mathbf{B} - \mathbf{C}|| = L_2$ (see

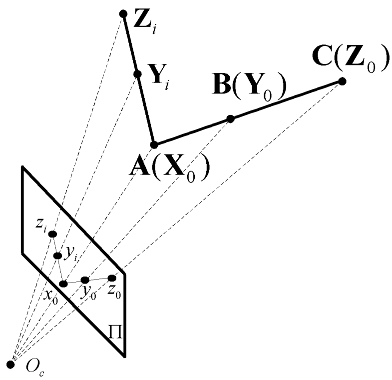


Fig. 1. Multi-camera system calibration with a 1D object

Fig.1). For the convenience of statement, the 1D calibration object is also called the line-segment (**ACB**) and denoted as $\mathbf{L_{ACB}}$.

2.3 Branch and Bound Theory

Branch and bound algorithms are methods for global optimization of non-convex problems [14]. They maintain a provable upper or(and) lower bound on the objective function and terminate with a ϵ-suboptimal for arbitrarily small ϵ.

For a multivariate, non-convex, scalar-valued objective function $f : \mathbf{R}^m \to \mathbf{R}$, to seek a global minimum over an m-dimensional rectangle Q_{init}, firstly introduce an auxiliary function $\Phi_{1b}(Q)$ which satisfies the following conditions for every region $Q \subset Q_{init}$.

(**R**1). $\Phi_{1b}(Q) \leq f_{min}(Q)$. Thus, the function $\Phi_{1b}(Q)$ compute a lower bound on $f_{min}(Q)$ of $f(x)$ for all $x \in Q$.

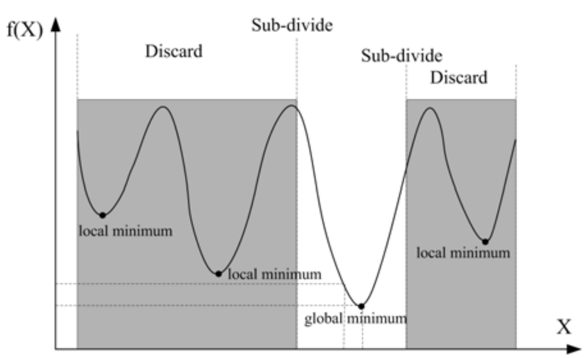


Fig. 2. The mechanism of global minimization with branch and bound algorithm for a univariate function

(**R**2). $\forall \epsilon > 0, \exists \delta > 0$ such that $\forall Q \subset Q_{init}, \mid Q \mid \leq \delta \Rightarrow f(x) - \Phi_{1b}(x) \leq \epsilon$, i.e., as $\mid Q \mid$, the length of the longest edge of a rectangle Q, goes to zero, the relaxation gap $f(x) - \Phi_{1b}(x)$ uniformly converges to zero.

Bounding refers to computing the value of $\Phi_{1b}(Q)$, while branching refers to choosing and subdividing a rectangle Q. Fig.2 illustrates the basic mechanism of a branch and bound routine for the case of a univariate function. Take the rectangle with the smallest minimum of Φ_{1b} as the most promising to contain the global minimum and pick it for refinement. Then subdivide the picked rectangle into k rectangles along the largest dimension. Repeat this process iteratively until the optimal solution is reached. The key problem of branch and bound methods is designing a proper bounding function $\Phi_{1b}(Q)$, which should be easy to estimate.

3 Global Optimal 1D Calibration Algorithm

The goal of camera 1D calibration is to compute the camera intrinsic matrix metric $\mathbf{K}$, with image points $\{\mathbf{a}_{ij}, \mathbf{b}_{ij}, \mathbf{c}_{ij} | i = 0, 1, 2, ..., m, j = 1, 2, ..., n\}$ of $\mathbf{L_{ACB}}$ under the jth rotation around the fixed point $\mathbf{A}$ (see Fig. 1). A camera system can be linearly calibrated with 1D calibration objects undertaken rotations [5]. Firstly express 3D coordinates with image points of markers $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{C}_i$ on 1D objects. Secondly construct equations with estimated 3D coordinates corresponding to image points and geometry constraints of 1D calibration objects. However the linear 1D calibration algorithm is susceptible to fail due to image noise. On the other hand, the conventional bundle adjustment optimization algorithm possibly converges to a local rather than a global minimum. So a 1D global optimal calibration algorithm using branch and bound framework is proposed to reduce the effect of image noises and improve the calibration accuracy.

3.1 Traditional Solution

As shown in Fig. 1, let the line segment $\mathbf{AC}$ rotates N times around the fixed point $\mathbf{A}$. Denote the coordinate of $\mathbf{A}$ by $\mathbf{X}_0$ under the camera coordinate system, and coordinates of $\mathbf{B}$ and $\mathbf{C}$ under the i^{th} rotation by $\mathbf{Y}_i$ and $\mathbf{Z}_i$, respectively, $(i = 0, 1, 2, \ldots, N)$, where $\mathbf{Y}_0$, $\mathbf{Z}_0$ is the coordinates of $\mathbf{B}$ and $\mathbf{C}$ at the initial position. Then, we have

$$\mathbf{Y}_i = \lambda_1 \mathbf{X}_0 + \lambda_2 \mathbf{Z}_i, \tag{2}$$

where $\lambda_1 = (L_1 - L_2)/L_1$ and $\lambda_2 = L_2/L_1$. Denote the image point of $\mathbf{X}_0$, $\mathbf{Y}_i$, $\mathbf{Z}_i$ by $\mathbf{x}_0$, $\mathbf{y}_i$, $\mathbf{z}_i$. With equation (1) under the camera coordinate system, we have

$$d_0^x \mathbf{x}_0 = \mathbf{K}\mathbf{X}_0, \quad d_i^y \mathbf{y}_i = \mathbf{K}\mathbf{Y}_i, \quad d_i^z \mathbf{z}_i = \mathbf{K}\mathbf{Z}_i. \tag{3}$$

Substitute equation (2) into $d_i^y \mathbf{y}_i = \mathbf{K}\mathbf{Y}_i$, then we have

$$d_i^y \mathbf{y}_i = d_0^x \lambda_1 \mathbf{x}_0 + d_i^z \lambda_2 \mathbf{z}_i, \quad i = 0, 1, 2, \cdots, N. \tag{4}$$

By some simple vector manipulation, we obtain the following constraints

$$d_i^y = d_0^x d'^y_i, \quad d_i^z = d_0^x d'^z_i, \tag{5}$$

where d'^y_i, d'^z_i is the relative depth of $\mathbf{Y}_i$, $\mathbf{Z}_i$ to $\mathbf{X}_0$:

$$d'^y_i = \frac{\lambda_1(\mathbf{y}_i \times \mathbf{z}_i)^T(\mathbf{x}_0 \times \mathbf{z}_i)}{\lambda_2(\mathbf{y}_i \times \mathbf{z}_i)^T(\mathbf{y}_i \times \mathbf{z}_i)}, \tag{6}$$

$$d'^z_i = \frac{\lambda_1(\mathbf{x}_0 \times \mathbf{y}_i)^T(\mathbf{z}_i \times \mathbf{y}_i)}{\lambda_2(\mathbf{z}_i \times \mathbf{y}_i)^T(\mathbf{z}_i \times \mathbf{y}_i)}. \tag{7}$$

With the geometry information of the 1D object, for each i ($i = 1, 2, \ldots, N$) we have the following constraint

$$L = \|\mathbf{Z}_i - \mathbf{X}_0\| = \|\mathbf{K}^{-1}(d_i^z \mathbf{z}_i - d_0^x \mathbf{x}_0)\|. \tag{8}$$

It can also be expressed as the following equation system,

$$(d'^z_i \mathbf{z}_i - \mathbf{x}_0)^T \varpi (d'^z_i \mathbf{z}_i - \mathbf{x}_0) = L^2 \quad i = 0, \cdots, N. \tag{9}$$

where $\varpi = (d_0^x)^2 \mathbf{K}^{-T}\mathbf{K}^{-1}$. Considering that

$$\begin{aligned} &\mathbf{m}_i^T \varpi \mathbf{m}_i = \mathbf{M}_i \Theta \\ &= [\mathbf{m}_{i1}^2 \quad 2\mathbf{m}_{i1}\mathbf{m}_{i2} \quad 2\mathbf{m}_{i1}\mathbf{m}_{i3} \quad \mathbf{m}_{i2}^2 \quad 2\mathbf{m}_{i2}\mathbf{m}_{i3} \quad \mathbf{m}_{i3}^2]\Theta \end{aligned} \tag{10}$$

with

$$\Theta = [\varpi_{11} \quad \varpi_{12} \quad \varpi_{13} \quad \varpi_{22} \quad \varpi_{23} \quad \varpi_{33}]^T \tag{11}$$

and $\mathbf{m}_i = (d'^z_i \mathbf{z}_i - \mathbf{x}_0)$, (9) can be rewritten as

$$\mathbf{M}_i \Theta = L^2 \quad i = 0, \cdots, N. \tag{12}$$

With $\mathbf{M} = [\mathbf{M}_0^T, \mathbf{M}_1^T, \cdots, \mathbf{M}_N^T]^T$, $\mathbf{L} = [L^2, L^2, \cdots, L^2]^T$, we can obtain Θ with the total (orthogonal) least squares method by sloving

$$\mathbf{M}\Theta = \mathbf{L}. \tag{13}$$

Once Θ is obtained with equation (11) we can compute ϖ, then obtain $d_0^x \mathbf{K}^{-1}$ by the Cholesky decomposition. Finally the inner matrix $\mathbf{K}$ is obtained.

3.2 Global Optimization with Branch and Bound

The above solution is obtained by minimizing algebraic distances which is susceptible to image noise. The conventional way to overcome this deficiency is resort to bundle adjustment, however it often traps in local optimal solution. We instead consider branch and bound algorithm outlined in this section to refine 1D calibration results.

In the linear 1D calibration, the main problem is to estimate the image of absolute quadratic (IAC)

$$\begin{aligned}\varpi_0 &= \mathbf{K}^{-T}\mathbf{K}^{-1}\\ &= \tfrac{1}{\alpha^2\beta^2}\begin{bmatrix} \beta^2 & -\gamma\beta & -u_0\beta^2+\gamma v_0\beta \\ -\gamma\beta & \alpha^2+\gamma^2 & \gamma u_0\alpha - v_0\alpha^2-\gamma^2 v_0 \\ -u_0\beta^2+\gamma v_0\beta & \gamma u_0\alpha - v_0\alpha^2-\gamma^2 v_0 & \alpha^2\beta^2+\alpha^2{v_0}^2+(\beta u_0-\gamma v_0)^2 \end{bmatrix}\end{aligned} \tag{14}$$

It can be posed as the following least squares problem:

$$\min_{\varpi_0}\sum_i (L_1^2-(\mathbf{Z}_i-\mathbf{X}_0)^T\varpi_0(\mathbf{Z}_i-\mathbf{X}_0))^2 \tag{15}$$

Considering equation (8), equation (15) can be rewritten as

$$\min_{\hat{\varpi}_0}\sum_i (L_1^2-k({d'}_i^z\mathbf{z}_i-\mathbf{x}_0)^T\hat{\varpi}_0({d'}_i^z\mathbf{z}_i-\mathbf{x}_0))^2 \tag{16}$$

where $k=(d_0^x)^2/\alpha^2/\beta^2$,

$$\hat{\varpi}_0 = \begin{bmatrix} \beta^2 & -\gamma\beta & -u_0\beta^2+\gamma v_0\beta \\ -\gamma\beta & \alpha^2+\gamma^2 & \gamma u_0\alpha - v_0\alpha^2-\gamma^2 v_0 \\ -u_0\beta^2+\gamma v_0\beta & \gamma u_0\alpha - v_0\alpha^2-\gamma^2 v_0 & \alpha^2\beta^2+\alpha^2{v_0}^2+(\beta u_0-\gamma v_0)^2 \end{bmatrix} \tag{17}$$

It is solved linearly by ignoring the positive semi-definiteness requirement on the IAC. For the cases where the linear solution does not yield a positive semi-definite IAC, the closest positive semi-definite matrix is estimated as a post-processing step by dropping the negative eigenvalues.

We take into account the positive semi-definiteness of the IAC in the optimization algorithm and search for the optimal IAC in the feasible domain. To this end, we pose the minimization problem:

$$\begin{aligned}&\min_{\hat{\varpi}_0}\sum_i (L_1^2-k({d'}_i^z\mathbf{z}_i-\mathbf{x}_0)^T\hat{\varpi}_0({d'}_i^z\mathbf{z}_i-\mathbf{x}_0))^2\\ &\textit{subject to}\quad \hat{\varpi}_0\succeq 0\\ &\qquad\qquad\qquad \hat{\varpi}_0\in\aleph\end{aligned}, \tag{18}$$

where $\hat{\varpi}_0\succeq 0$ describes the positive semi-definiteness, $\aleph$ is some initial convex region which can be given by estimated camera matrix $\bar{\mathbf{K}}$ with original linear 1D calibration. To improve the numerical conditioning of this method, the matrix

$$\mathbf{V}=\frac{1}{2}\begin{bmatrix} \sqrt{w^2+h^2} & 0 & w \\ 0 & \sqrt{w^2+h^2} & h \\ 0 & 0 & 2 \end{bmatrix} \tag{19}$$

is used to normalize $\hat{\mathbf{K}}$ by $\mathbf{K}=\mathbf{V}^{-1}\bar{\mathbf{K}}$, where w and h are respectively the width and height of images.

Assume that the domain $\aleph$ is given in the form of bounds $[l_{jk}, r_{jk}]$ on the five unknown symmetric entries $\hat{\varpi}_{0jk}$ of $\hat{\varpi}_0$. Then the above optimization problem can be stated as

$$\begin{array}{ll} \min\limits_{\hat{\varpi}_0} \sum\limits_i (L_1^2 - k(d'^z_i \mathbf{z}_i - \mathbf{x}_0)^T \hat{\varpi}_0 (d'^z_i \mathbf{z}_i - \mathbf{x}_0))^2 & \\ subject \quad to \quad \hat{\varpi}_0 \succeq 0 & . \\ \qquad\qquad\qquad l_{jk} \leq \hat{\varpi}_{0jk} \leq r_{jk} & \end{array} \tag{20}$$

Then introduce a new matrix $\nu = k\hat{\varpi}_0$. The objective function of the optimization problem (20) can be expressed as $\min\limits_{\hat{\varpi}_0} \sum\limits_i (L_1^2 - (d'^z_i \mathbf{z}_i - \mathbf{x}_0)^T \nu (d'^z_i \mathbf{z}_i - \mathbf{x}_0))^2$. With the bilinear equality constraint $(d'^z_i \mathbf{z}_i - \mathbf{x}_0)^T \nu (d'^z_i \mathbf{z}_i - \mathbf{x}_0) = L_1^2$, the bounds of variable of k can be given by simply inverting the bounds on $\hat{\varpi}_0$. The convex relaxation of the above optimization problem can be stated as

$$\begin{array}{ll} \min\limits_{\hat{\varpi}_0, k, \nu} \sum\limits_i (L_1^2 - (d'^z_i \mathbf{z}_i - \mathbf{x}_0)^T \nu (d'^z_i \mathbf{z}_i - \mathbf{x}_0))^2 & \\ subject \quad to \quad \nu_{jk} \leq k_u \hat{\varpi}_{jk} + l_{jk} k - k_u l_{jk} & \\ \qquad\qquad\qquad \nu_{jk} \leq k_l \hat{\varpi}_{jk} + u_{jk} k - k_l u_{jk} & \\ \qquad\qquad\qquad \nu_{jk} \geq k_l \hat{\varpi}_{jk} + l_{jk} k - k_l l_{jk} & . \\ \qquad\qquad\qquad \nu_{jk} \geq k_u \hat{\varpi}_{jk} + u_{jk} k - k_u u_{jk} & \\ \qquad\qquad\qquad \hat{\varpi}_0 \succeq 0 & \\ \qquad\qquad\qquad l_{jk} \leq \hat{\varpi}_{0jk} \leq r_{jk} & \\ \qquad\qquad\qquad k_l \leq k \leq k_r & \end{array} \tag{21}$$

The objective function of the optimization problem (21) is convex quadratic. The constraint set includes linear inequalities and a positive semi-definiteness constraint. Such problem can be efficiently solved using interior point methods which can obtain their global optimum and have a number of software packages for implementation. In our implementation, SeDuMi [19] is used.

Once $\hat{\varpi}_0$ is obtained, the normalized camera matrix $\hat{\mathbf{K}}$ can be computed by equation (17). Finally the camera matrix can be obtained by de-normalization with $\mathbf{K} = \mathbf{V}\hat{\mathbf{K}}$.

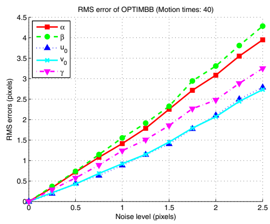
4 Experiments

To validate the proposed algorithm, computer simulated data and real image data experiments have been performed. Some results are reported in this section.

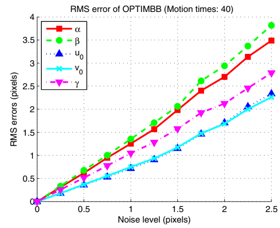
4.1 Simulated Data

In this experiment, the intrinsic parameters of the simulated camera are $\alpha = 500$, $\beta = 500$, $u_0 = 512$, $v_0 = 512$ and $\gamma = 0.0$. The distance between two markers $\mathbf{A}$ and $\mathbf{C}$ on the 1D object is $L_1 = 20$, the distance between two markers $\mathbf{B}$ and $\mathbf{C}$ is $L_2 = 10$. The initial positions of three markers are $\mathbf{A} = (0, 0, 25)^T$, $\mathbf{B} = (5\sqrt{2}, 0, 25 + 5\sqrt{2})^T$, $\mathbf{C} = (10\sqrt{2}, 0, 25 + 10\sqrt{2})^T$ respectively. Rotate the 1D object N times with randomly selected angle and axis around fixed point

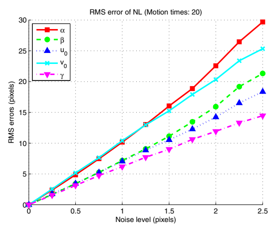
A, then $2N + 3$ space points can be obtained which include three points on the initial position. Assumed that the camera coordinate system coincides with the world coordinate system, image points of these markers can be obtained by equation (1).



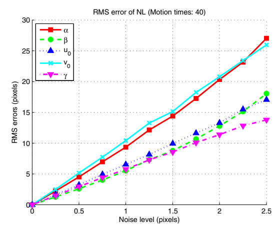
(a) OPTIMBB with 20 rotations



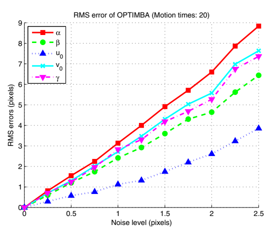
(b) OPTIMBB with 40 rotations



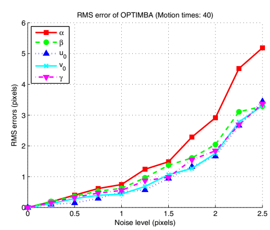
(c) NL with 20 rotations



(d) NL with 40 rotations



(e) OPTIMBA with 20 rotations



(f) OPTIMBA with 40 rotations

Fig. 3. Experimental results of the proposed optimal, NL and conventional bundle adjustment 1D calibration with 20 and 40 rotations of 1D object

Add Gaussian noise with mean 0 and standard derivations σ to image points. The noise level σ is varied from 0 to 2.5 pixels with a step of 0.25 pixels. Here we use three methods, the proposed optimal algorithm based on branch and bound framework (OPTIMBB), the normalized linear algorithm (NL) and the conventional optimal algorithm based on bundle adjustment (OPTIMBA), to calibrate the camera. In order to provide more statistically meaningful results, for each noise level we perform 1000 trials. The RMS errors of estimated inner parameters relative to the ground truth are computed to evaluate the performance. Results are shown in Figure 3.

We can see in Figure 3, errors of the proposed OPTIMBB algorithm are less than that of the normalized linear algorithm and the conventional OPTIMBA algorithm. Errors of the proposed OPTIMBB algorithm increase almost linearly with the noise level while the conventional OPTIMBA algorithm does not linearly increase with noise level due to trapping into local minima.

4.2 Real Images Data

For the experiment with real data, distances between two adjacent markers on the 1D object are 15cm. Rotate the 1D object 15 times around one fixed marker. Then use measurements of makers' image to calibrate the camera with the OPTIMBB, NL and OPTIMBA 1D calibration. We also perform 2D calibration and take its results as the ground truth. Calibration results are shown in Table 1. We can see that the performance of the proposed OPTIMBB algorithm is better than the other algorithm.

Table 1. Experimental results with real image data

	α	β	γ	u_0	v_0
OPTIMBB	2097	2080	-0.4	764	525
NL	1923	1938	-1.8	806	552
OPTIMBA	2034	1988	1.9	762	537
2D	2075	2078	0	757	514

5 Conclusions

In this paper, we have investigated to improve the accuracy of 1D calibration. A global optimal algorithm based on branch and bound framework is proposed. By constructing tight convex relaxations of the objective functions and minimizing in a branch and bound optimization framework, the optimal solution, which can provide ε-optimality, is obtained. Experiments prove the validity of the proposed method.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (Grant No.60872127).

References

1. Abdel-Aziz, Y.I., Karara, H.M.: Direct linear transformation from comparator coordinates into object space coordinates. In: ASP Symposium on Colse-Range Photogrammetry, Virginia, USA (1971)
2. Tsai, R.: An efficient and accurate camera calibration technique for 3d machine vision. In: Proceedings of the CVPR 1986, Miami Beach, USA, pp. 364–374 (1986)
3. Zhang, Z.Y.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the ICCV 1999, Kerkya, Greece, pp. 666–673 (1999)
4. Meng, X.Q., Li, H., Hu, Z.Y.: A new easy camera calibration technique based on circular points. In: Proceedings of the BMVC 2000, Bristol, UK, pp. 496–505 (2000)
5. Zhang, Z.Y.: Camera calibration with one-dimensional objects. IEEE Trans. Pattern Anal. Mach. Intell. 26(7), 892–899 (2004)
6. Hammarstedt, P., Sturm, P., Heyden, A.: Degenerate cases and closed-form solutions for camera calibration with one-dimensional objects. In: Proceedings of the ICCV 2005, Beijing, China, pp. 317–324 (2005)
7. Wu, F.C., Hu, Z.Y., Zhu, H.J.: Camera calibration with moving one-dimensional objects. Pattern Recognition 38(5), 755–765 (2005)
8. Wang, L., Wu, F.C., Hu, Z.Y.: Multi-camera calibration with one-dimensional object under general motions. In: Proc. IEEE Int. Conf. Comput. Vis., pp.1–7 (2007)
9. Franca, J., Stemmer, M.R., et al.: Revisiting Zhang's 1D calibration algorithm. Pattern Recognition 43(3), 1180–1187 (2010)
10. Maybank, S.J., Faugeras, O.D.: A theory of self-calibration of a moving camera. Int. J. Comput. Vision 8(2), 123–152 (1992)
11. Pollefeys, M., Gool, L.V., Oosterlinck, A.: The modulus constraint: a new constraint for self-calibration. In: Proceedings of the ICPR 1996, Vienna, Austria, pp. 31–42 (1996)
12. Triggs, B.: Auto-calibration and the absolute quadric. In: Proceedings of the CVPR 1997, Puerto Rico, USA, pp. 609–614 (1997)
13. Triggs, B.: Autocalibration from planar scenes. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 89–105. Springer, Heidelberg (1998)
14. Land, A.H., Doig, A.G.: An automatic method of solving discrete programming prolems. Econometrica 28(3), 497–520 (1960)
15. Chandraker, M., Agarwal, S., et al.: Globally optimal algorithms for stratified autocalibration. Int. J. Comput. Vision 90(2), 236–254 (2010)
16. Hartley, R., Zisserman, A.: Mmultiple view geometry in computer vision. Cambridge University Press, Cambridge (2000)
17. More, J.J.: The levenberg-marquardt algorithm, implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis. Springer, Heidelberg (1977)
18. Rother, C.: Multi-view reconstruction and camera recovery using a real or virtual reference plane, PhD thesis, Computational and Active Perception Laboratory, Kungl Tekniska Högskolan (2003)
19. Sturm, J.: Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. Optimization Methods and Software 11(12), 625–653 (1999)

LSCM Based Non-rigid Registration for Craniofacial Surfaces

Wenkui Xie, Fuqing Duan, Qingqiong Deng,
Mingquan Zhou, ZhongkeWu, and Liang Chang

College of Information Science and Technology,
Beijing Normal University, Beijing 100875, P.R. China

Abstract. The Least Squares Conformal Maps (LSCM) is an approximation of the conformal mapping in the least–squares sense, and it can map the corresponding feature points on two 3D surfaces into the same 2D location. This paper proposes a non-rigid registration method for craniofacial surfaces based on LSCM parameterization. Firstly, craniofacial surfaces are normalized in pose and scale by using a unified coordinate system. Secondly, by pinning six landmarks, which include the outer corners of the eyes, two corners of the mouth, two side points of the nose wing, each craniofacial surface is mapped into a nearly equal 2D domain by using LSCM. Finally, an iso-parameter mesh of each craniofacial surface can be obtained by 2D to 3D mapping, which establishes a unique correspondence among different craniofacial surfaces. To evaluate the proposed method, the target surface is deformed into the reference surface using TPS algorithm with dense correspondences being control points, and then the sum of the distance between two correspondence point sets are computed, and vice versa. According to the average distance, the proposed method is compared with ICP and a TPS based method. The comparison shows that the proposed approach is more accurate and effective.

Keywords: non-rigid registration, LCSM, craniofacial surface.

1 Introduction

Non-rigid registration of 3D data set is an important and challenging issue for many applications such as atlas matching, shape retrieval, 3D object recognition etc [1,2]. The registration for craniofacial surfaces is to find an optimal transformation from one craniofacial surface to the other surface, so that each point of one craniofacial surface maps to its corresponding point in another surface which stems from the same physical point. In fact, the registration for craniofacial surfaces is not a well-defined problem. Except for some distinct points such as the tip of nose and the corners of mouth and so on, it is difficult to define exact correspondence between different craniofacial surfaces, especially for the points on smooth regions such as the cheeks and forehead.

Up to now, there are many researches on the 3D registration. The iterative closest point (ICP) algorithm [3] is a classical technique used in 3D registration. It iteratively searches the closest corresponding points in two data sets and optimized rigid

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 670–676, 2011.
© Springer-Verlag Berlin Heidelberg 2011

transformation to minimize the distances between these closest points. Since the original ICP algorithm has a heavy time cost and need a good pre-alignment, various improved ICP methods [4] have been proposed to enhance the registration accuracy or convergence stability of ICP. However, ICP is not suitable for alignments with large non-rigid deformation. The dominant non-rigid deformation is Thin Plate Spline (TPS)-based [5,6]. Hutton et al [7] proposed a dense 3D face model registration based on TPS. They manually pick up 9 feature points as TPS controlling points on 3D faces in the training set. However, these 9 feature points are inadequate for describing the variation of the 3D facial shapes. Schneider and Eisert [8] propose an automatic registration method by combining ICP and TPS. In this method, landmarks are first automatically defined using the ICP scheme with a re-weighted error function, and then a TPS deformation is computed based on the landmark correspondences. Hu et al. [9] propose a similar method, but they define the landmarks of TPS by first randomly selecting some points on one face, and then finding their correspondences on the other face through an iterative closest point searching strategy. However, closest point pairs may not always physically correspond to each other.

In this work, we propose a registration method for 3D craniofacial surfaces based on Least squares conformal maps parameterization (LSCM). LSCM is an approximation of the conformal mapping in the least–squares sense, and it can map the corresponding feature points on two 3D surfaces into the same 2D location, and can map a 3D surface to a 2D plane in a continuous manner with minimized local angle distortion. By pinning six corresponding landmarks on each craniofacial surface, all craniofacial surfaces are mapped into nearly equal 2D parameterizations using LSCM, that is, the corresponding points are mapped onto nearly equal 2D parameters. Compared with ICP and a TPS based method, the proposed approach is more accurate and effective.

Section 2 introduces the LSCM. Section 3 describes the proposed registration method for craniofacial surfaces. Experimental results are reported in Section 4, and followed are some conclusions in Section 5.

2 Least Squares Conformal Maps

In this section, we briefly introduce the notion of the LCSM (see [10] for details).

It can be proofed by Riemann's theorem that any surface homeomorphic to a disc can be parameterized to a 2D planar domain by a conformal mapping, which is one to one, onto, and angle preserving [11,12]. The mapping can be uniquely determined by any two points on the surface. However, the mapping usually is unreliable, since there is noise in the discrete data we obtained. In order to better handle the errors caused by noise and the inaccuracy of locating feature points, the LSCM introduces additional constraints in the least squares sense.

Consider a discrete 3D surface triangle mesh S and a smooth target mapping $U : S \rightarrow (u,v)$. U is conformal on S if and only if the following Cauchy-Riemann equation holds true on the whole of S.

$$\frac{\partial U}{\partial x} + i\frac{\partial U}{\partial y} = 0 \tag{1}$$

Since this conformal condition cannot be strictly satisfied on the whole triangulated surface S, the conformal map is constructed in the least squares sense with the constraint that the mapping U is linear on each triangle of the mesh surface S. Then the conformal map is constructed in the least-squares sense:

$$MinC(S) = \sum_{d \in S} \left| \frac{\partial U}{\partial x} + i \frac{\partial U}{\partial y} \right|^2 A(d) \tag{2}$$

where d is a triangle on the mesh S, and $A(d)$ is the area of the triangle d . In solving this minimization problem, we can add more correspondences as additional constraints by pinning multiple points with a priori values. Thus, we can map a 3D surface to a 2D domain with multiple correspondences as constraints by using the LSCM technique.

Compared to conformal maps, LSCM has the following advantages [12]: LSCM can map a 3D shape to a 2D domain in a continuous manner with minimized local angle distortion; LSCM can handle missing boundaries and occlusion; LSCM is independent of mesh resolution, LSCM can be linearly solved; LSCM can better handle noise from the feature point detection due to multiple feature constraints.

3 The Registration Algorithm for Craniofacial Surfaces

3.1 Normalization of Craniofacial Surfaces

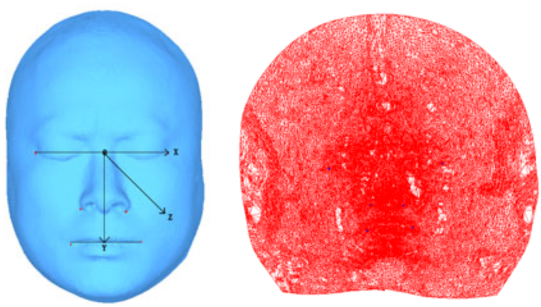
Since different craniofacial data can be obtained by different 3D data acquisition, and the poses and the scales of the craniofacial surfaces are various, we need to perform a pose and scale normalization before registration. According to human physiological characteristics, the four landmarks, which include the two outer corners of eyes and two corners of the mouth, are coplanar. We build a unified coordinate system as follows: The midpoint of the two landmarks of two eyes is the origin of the coordinate system; The direction from the coordinate origin to the right landmark of the eyes is the x-direction; The normal of the plane is the z-direction; Then, y-direction is determined according to the right-hand rule; Finally, the metric unit of the coordinate system is defined according to the distance from the coordinate origin to the right landmark of the eyes. Figure.1a shows the unified coordinate system. When transformed to this unified coordinate system, each craniofacial surface can be adjusted to a same pose and a same scale.

3.2 Registration Based on LCSM

A craniofacial surface is approximately homeomorphic to a disc, so we can use LSCM to map a craniofacial surface to a 2D domain. When adding additional feature constraints that the corresponding anatomy features on any craniofacial surface should be mapped onto the same 2D locations, we expect that corresponding points on different craniofacial surfaces have nearly the same 2D values. Then from 2D correspondences, we can obtain the 3D point matching.

To generate feature constraints in LSCM, we have to pin some feature points. As we know, except for some distinct points such as the tip of nose and the corners of mouth and so on, it is difficult to find the exact anatomy correspondence between different craniofacial surfaces. In addition, as Figure.1 shows, the six landmarks in a craniofacial surface including the outer corners of the two eyes, two corners of the mouth, two side points of the nose wing, control the main feature region of a craniofacial surface. So we choose the six landmarks to generate the feature constraints in LSCM. However, improper constraints will generate bad mapping, and how to assign proper prior 2D values to these six landmarks is not easy. We firstly map one reference surface into 2D domain using LSCM by pinning the two landmarks of eyes, where the two landmarks are assigned (-1, 0) and (1, 0) respectively. The 2D domain of the reference surface is shown in Figure.1b, where the 2D locations of the six landmarks are also labeled in Figure.1b. Table.1 shows the 2D values of the six landmarks. Then any craniofacial surface can be mapped to a 2D domain by LSCM with pinning the six landmarks according to the obtained 2D values, Fig.2 shows another craniofacial surface and the 2D domain mapped by LSCM with pinning the six landmarks.

Since the LSCM subjects to the constraint that the mapping is linear on each triangle of the mesh surface, the iso-parametric mesh of a craniofacial surface can be obtained by linear interpolation, and a 3D point correspondence is formed according to the iso-parametric mesh. Fig.2c shows the registration result between two craniofacial surfaces in Fig.1 and Fig.2.

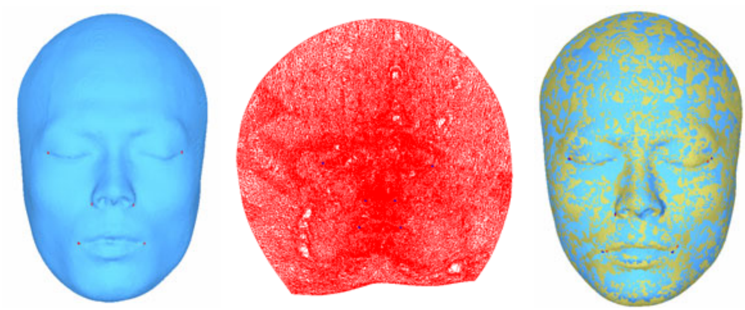


a b

Fig. 1. a)The unified coordinate system and the six landmarks, b)The 2D domain by LSCM

Table 1. The 2D values of the six landmarks

2D values of landmarks	1	2	3	4	5	6
u	-1.0	1.0	-0.29	0.28	-0.37	0.43
v	0.0	0.0	-0.68	-0.71	-1.13	-1.11



a b c

Fig. 2. a)A craniofacial surface, b) The 2D domain by LSCM with pinning six feature points, c)The registration result

4 Result

In this section, we do some experiments to evaluate our algorithm. We also use the two craniofacial surfaces in Section3, which are acquired by reconstructing from real CT images. The craniofacial surface in Fig.1 has 28610 vertices and 55917 triangles, and the one in Fig.2 has 28421 points and 55296 triangles. It is obviously has much differences between them in shape and geometry feature. In experiments, since the back of a head has less recognition features, a vertical cut is performed manually to retain the face part before the ears. The platform of the experiments is Dell precision 390 workstation with Intel(R) Core(TM) 2.39GHz and 3GB memory. The registration result in Fig. 2c shows that the faces can match well with each other, which indicates the proposed method is effective. In following, we compare our method with ICP and TPS.

Since the point number of the models is large, it will take a long time for registration using all points. So, we do random sampling to select points for the ICP registration, which distribute uniformly on the surfaces, and the distances among these points are forced being greater than a given threshold. In the experiment, 2500 points are select on each face. Fig. 3 shows the selected points, including most of characters on the reference face. The 2500 points are used to compute the transform between two surfaces in ICP. The result using ICP is shown in Fig. 4a. Using these 2500 point correspondences obtained by ICP as controlling points, the reference craniofacial is deformed to target craniofacial by TPS, and the result using TPS is shown in Fig. 4b. The result using the proposed method is also shown in Fig. 4c. To make the result clearer, frontal faces are displayed with solid, and side faces use triangles. We can see that the quality of ICP registration is the worst, and most of the points match with each other perfectly by the proposed method, which is better than the TPS method.

Furthermore, in order to evaluate the three methods quantitatively, we compute the mean distance between dense correspondences for ICP and TPS respectively, and for the proposed method, the target surface is deformed into the reference surface using TPS algorithm with dense correspondences being control points, and then the distance sum between two correspondence point sets are computed, and vice versa. The mean distances for the three methods are shown in Table 2. From the table, we can see the non-rigid method based on LSCM performs much better then ICP and TPS.

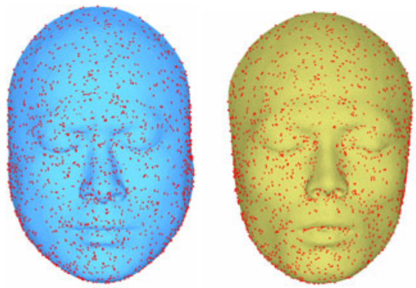


Fig. 3. The random sampled points on the reference face and the target face for TPS

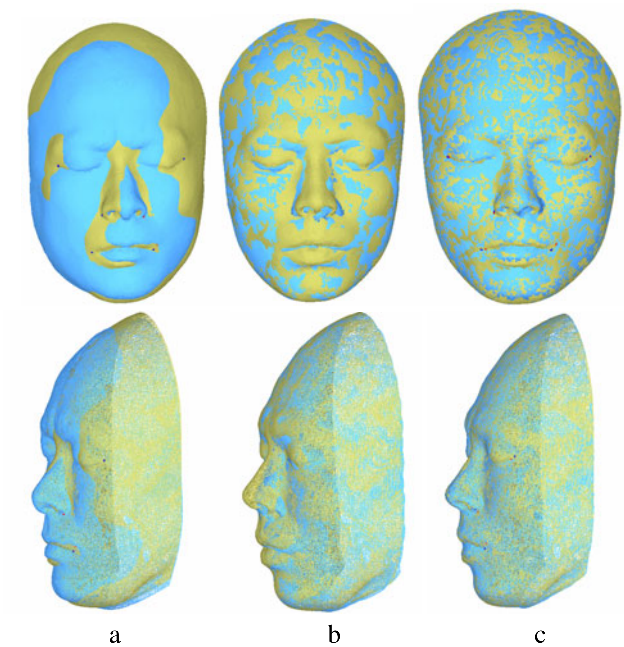


Fig. 4. a)The alignment of the two faces after the ICP transformation, b) The alignment of the two faces after the TPS transformation, c)The alignment of the two faces after our method

Table 2. The result of the criteria of ICP and Our method

Methods	ICP	TPS	Our method
Mean distance(mm)	3.01	0.71	0.49

5 Conclusions

Non-rigid registration of 3D data set is an important and challenging issue for many applications such as atlas matching; shape retrieval, 3D object recognition etc. This paper proposes a non-rigid registration method for craniofacial surfaces based on LSCM parameterization. Since LSCM can map the corresponding feature points on two 3D surfaces into the same 2D location, each craniofacial surface can be mapped into a nearly equal 2D domain by using LSCM. Then from 2D correspondences, we can obtain the 3D point matching. The proposed method is compared with ICP and a TPS based method. The comparison shows that the proposed approach is more accurate and effective. The future work we will do is to refine the registration result in 2D domain.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China No.60736008 and No.61003133, National High-Tech Research and Development 863 Plan of China under Grant No. 2008AA01Z301.

References

1. Antoine Maintz, J.B., Viergever, M.A.: A survey of medical image registration. Medical Image Analysisi 2(1) (March 1998)
2. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: IJCAI 1981, pp. 674–679 (1981)
3. Besl, P., McKay, N.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
4. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: The Third International Conference on 3D Digital Imaging and Modeling, pp. 145–152 (2001)
5. Bookstein, F.L.: Principal Warps: Thin-plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
6. Brown, B., Rusinkiewicz, S.: Nonrigid Range-Scan Alignment Using Thin-Plate Splines. In: Proc. Symp. Three-Dimensional Data Processing, Visualization, and Transmission (2004)
7. Hutton, T.J., Buxton, B.F., Hammond, P.: Automated Registration of 3D Faces using Dense Surface Models. In: The 2003 British Machine Vision Conference, pp. 439–448 (2003)
8. Schneider, D., Eisert, P.: Algorithms for Automatic and Robust Registration of 3D Head Scans. Journal of Virtual Reality and Broadcasting (2010)
9. Hu, Y., Zhou, M., Wu, Z.: A Dense Point-to-Point Alignment Method for Realistic 3D FaceMorphing and Animation. International Journal of Computer Games Technology, 1–9 (2009)
10. Levy, B., Petitjean, S., Ray, N., Maillot, J.: Least Squares Conformal Maps for Automatic Texture Atlas Generation. In: Proc. ACM SIGGRAPH Conf., pp. 362–371 (2002)
11. Haker, S., Angenent, S., Tannenbaum, A., Kikinis, R., Sapiro, G., Halle, M.: Conformal Surface Parameterization for Texture Mapping. IEEE Trans. Visualization and Computer Graphics 6, 181–189 (2000)
12. Wang, S., Wang, Y., Jin, M., Gu, X.D., Samaras, D.: Conformal Geometry and Its Applications on 3D Shape Matching, Recognition, and Stitching. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(7), 1209–1220 (2007)

High-Quality Fast Image Upsampling Algorithm Based on CUDA

Qingqing Xu[1], Xin Zheng[1], and Jie Chen[2]

[1] Image Processing and Pattern Recognition Laboratory,
College of Information Science and Technology,
Beijing Normal University, Beijing 100875, P.R. China
[2] Information center, CITIC Guoan Group, Beijing, P.R. China
zhengxin@bnu.edu.cn

Abstract. Although many upsampling methods have been proposed, but no method can get result images with satisfactory quality in real-time. In this paper, we propose a CUDA based image upsampling algorithm, which can generate sharp edges with reduced grid-related artifacts efficiently. By analyzing existing method, we find chock points which confine the efficiency of the algorithm mostly and use CUDA to accelerate our algorithm and improve the implementation model of the algorithm. In this way we not only guarantee the quality of the result image, but also realize the purpose of a real-time human-computer interaction. Experimental results show that our method can get high-quality upsampled images efficiently.

Keywords: image upsampling, CUDA, large-scale data parallel computing, image interpolation, super-resolution.

1 Introduction

As one of the most elementary image operations, image resizing or resampling has been used for many purposes and is supported by almost all image editing softwares. A proper linear pre-filtering can obtain satisfactory downsampled images. But for upsampling it is not the same case. Because of lacking the needed information during upsampling, the result images usually lack small-scale texture-related features and the sharp edges become blurry, original pixel grids are still noticeable. Although many interpolation-based upsampling methods[1-5] have been proposed, the quality of result images is not satisfactory.

In order to avoid the blurred areas in sharp edges, another kind of upsampling method is proposed by using edge detecting. The typical upsampling algorithms such as POCS[6] based on adaptive image magnification, upsampling via imposed edges statistics[9] and so on. These kinds of algorithms considered not only pixel connection but also sharp edges, therefore the sampling effects were satisfactory. But all these methods used quite complex edges detecting computation, therefore their speed were not suitable to use in real-time texture mapping.

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 677–683, 2011.
© Springer-Verlag Berlin Heidelberg 2011

In this paper, we propose a high-quality fast image upsampling algorithm based on CUDA, which can generate sharp edges with reduced grid-related artifacts efficiently. As the same with upsampling method via imposed edges statistics, our method is also based on a statistical edge dependency relating certain edge features of two different resolutions, which is generically exhibited by real-world images. A special edge reliance is calculated in the preprocessing, which will be used as a known condition to reach the goal of improving the image resolution. And in this process the intensity of the image must also be preserved in order to achieve the association which can assure the content of output and input images are the same while the original resolution image is downsampled.

Furthermore, throughout the whole upsampling process, the complex operations between matrices often confine the efficiency of the algorithm, therefore we use CUDA[10] to accelerate our algorithm and improve the implementation model of the algorithm. In this way we not only guarantee the quality of the images upsampled, but also realize the purpose of a real-time human-computer interaction when upsampling.

The rest of our paper is organized as follows. In section 2, we discuss the related work in upsampling method and CUDA[10]. In section 3, we discuss our algorithm in detail. Implementations and experiments are discussed in section 4. In section 5, we describe our conclusion.

2 Related Work

Upsampling is an important method for image processing and is widely used in different procedures of image editing. There are many kinds of different upsampling methods, such as Classical Methods, Weights Adapting Methods, Storing Additional Data, Edge Statistics and so on.

Classical Methods include Nearest-Neighbor, Bilinear, Bicubic and so on. They[1-5] are very easy and popular used in the contemporary image processing commercial softwares. But these methods always rely on the assumption that the image data is either spatially smooth or band-limited. Because of this, these methods may have bad visual effects during the image processing. Visual artifacts such as ringing, aliasing, blocking and blurring are obviously.

On the other hand, many people put the Weights Adapting Methods forward in order to avoiding blurring, ringing and other artifacts. One of these methods is POCS[6] based on adaptive image magnification. It was proposed by Ratakonda and Ahuja in 1998. The selective interpolation is implemented by using an iterative Projection Onto Convex Set (POCS) scheme. There are three basic steps in this method. Firstly, finding the edges of the input image. Secondly, obtaining the initial image. Finally, use POCS based on iterative algorithm method to get good upsampling images. Another Weights Adapting Method is Image Interpolation by using pixelleveldata-dependent triangulation. It was proposed by Su and Willis[7]. This method includes series of interpolation weights which are adjusted locally by choosing three out of the four nearest pixels. By using this method, we can reduce the number of variables that are averaged. This choice forms a noticeable block-like eggect, showing strong continuity along one of the two diagonals.

The Storing Additional Data methods also have many different methods. Tumblin and Choudhury[8] store additional image data in the form of discontinuity graphs and avoid the averaging pixels across boundaries. Raanan Fattal [9] uses Edge Statistics to upsample images. Although Fattal's method can get result images with high quality, it has a fatal weakness. Because of the accurately calculation, the speed of his method is very slow. So it can't be used in real time image processing.

At the same time, CUDA is NVIDIA's parallel computing architecture that enable dramatic increases in computing performance by harnessing the power of the GPU(Graphics Processing Unit)[10]. With millions of CUDA-enabled GPUs sold to date, software developers, scientists and researchers are finding broad-ranging uses for CUDA, including image and video processing, computational biology and chemistry, fluid dynamics simulation, CT image reconstruction, seismic analysis, ray tracing, and much more.

In our paper, we propose a method and use CUDA to accelerate the speed of the processing procedure which can get good results and can be used in real time image processing. So our method have an extensive use in many different fields which need to use image upsampling.

3 Upsampling Algorithm Based on CUDA

As a result of our research on upsampling methods, we know that the method of Image Upsampling via Imposed Edge Statistics produces the best result images. This algorithm was proposed by Raanan Fattal in 2007. It uses the edge statistics information to let the edges of upsampled images become clear without ringing, aliasing, blocking and blurring. The basic idea of Fattal's method can be described as Fig.1. By the guiding of this idea, we can clearly know the goal of our method. We'll improve Fattal's algorithm and accelerate the algorithm by using CUDA.

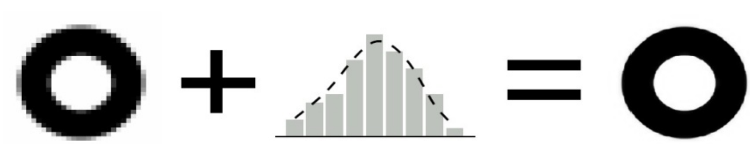


Fig. 1. Sharp upsampled image resulting from a low-resolution image plus edge statistics

3.1 Using Gradients Model to Get Good Upsampling Results

We use the changes in one direction of a gradient to judge the edges of an image. We sum the gradients of pixels along one edge and solve the average of the sum gradients. Thus, we can know the variance of the gradients. Use the average and the variance, we can construct a Gaussian distribution. Then by using this Gaussian distribution we construct a Gibbs distribution. The Gibbs distribution is a method for reduce the difficulty of the calculation of Gaussian distribution. In the condition of weak assumptions, the error rate of Gibbs distribution is only twice of the best classification(Bayes). So Gibbs distribution is a good choice for reducing calculated amount.

Having known the above conclusion, we will discuss the details in the following. The whole algorithm flow chart shows in Fig.2.

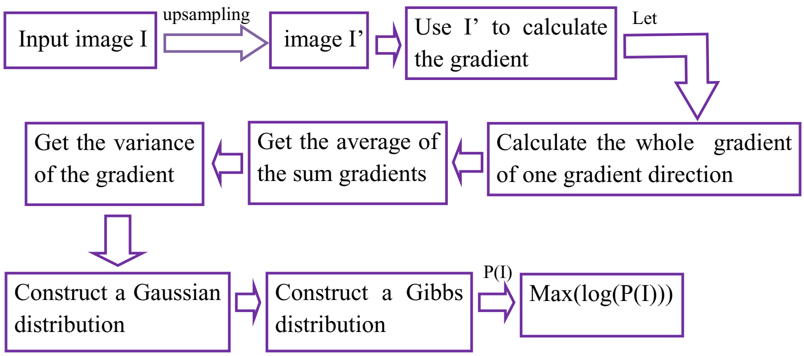


Fig. 2. The whole algorithm flow chart

3.2 Using CUBLAS to Accelerate the Algorithm

By analyzing the above processing procedure, we find that all steps with matrix operations are time cost process. While matrix operation is very suitable to be accelerated by using CUDA, we propose our CUDA based upsampling algorithm. By comparing 4 ways of matrix operation with CUDA: Without Shared Memory, With Shared Memory and Blocks, With Registers, Using CUBLAS, we use the most efficient method named CUBLAS to accelerate the computational process. We can see the whole procedure in Fig.3.

CUBLAS[10] is a BLAS(Basic Linear Algebra Subprograms) library ported to CUDA, which enables the use of fast computing by GPUs without direct operation of the CUDA drivers. This library uses on GPU per a process.

4 Result

We do experiments on a Intel(R) Core(TM) i3 CPU, 3.07GHz, 2G RAM computer. The GPU is NVIDIA GeForce 310 and coded by using C++. We will discuss the result of our method from three aspects: comparison in run-time, the result images of our method and the usage in large images.

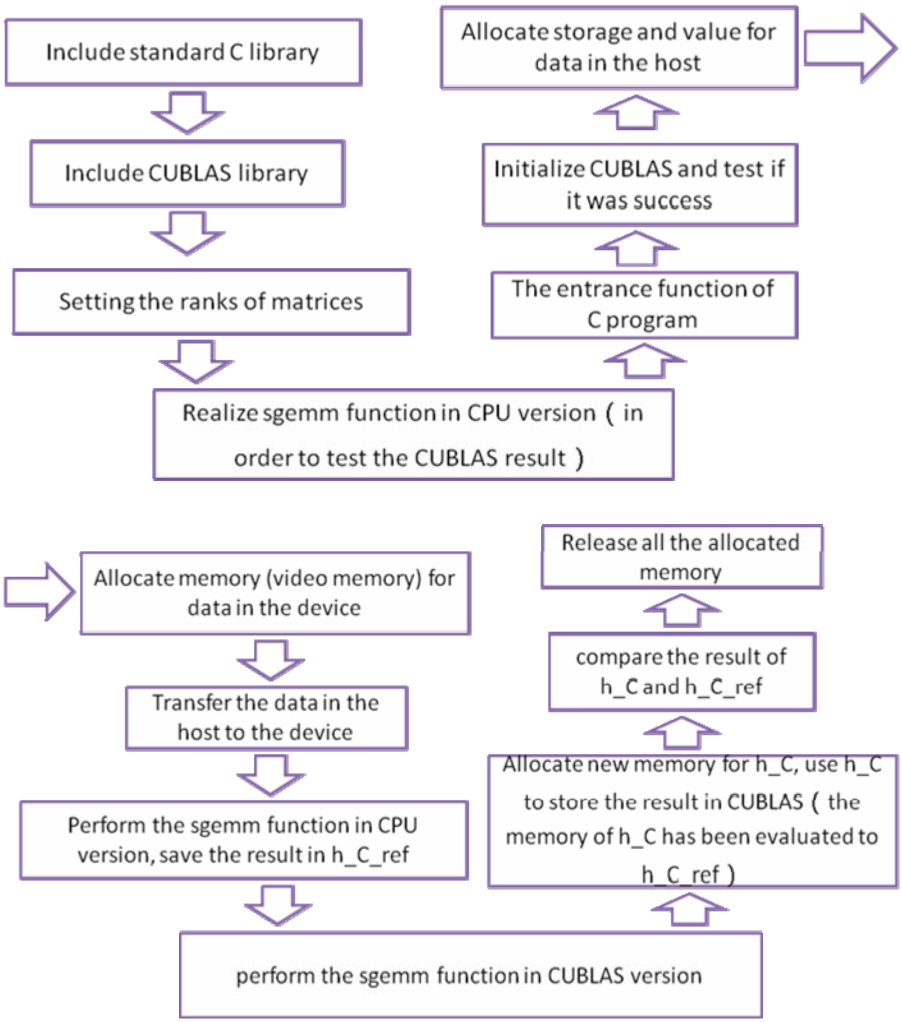


Fig. 3. The flow chart of the calculation of matrix by using CUBLAS

4.1 Comparison in Run-Time

By using the CUBLAS library to accelerate our algorithm, in the experiment, our method has a very good performance. Compared to Fattal's algorithm, our method is about three times faster. We can see the comparison data in Table 1.

We use different color images of size 128×128, 512×512 and 1024×1024 pixels in these experiments. In Table 1. we can see that our method has reduced the run-time obviously.

Table 1. The run-time comparation

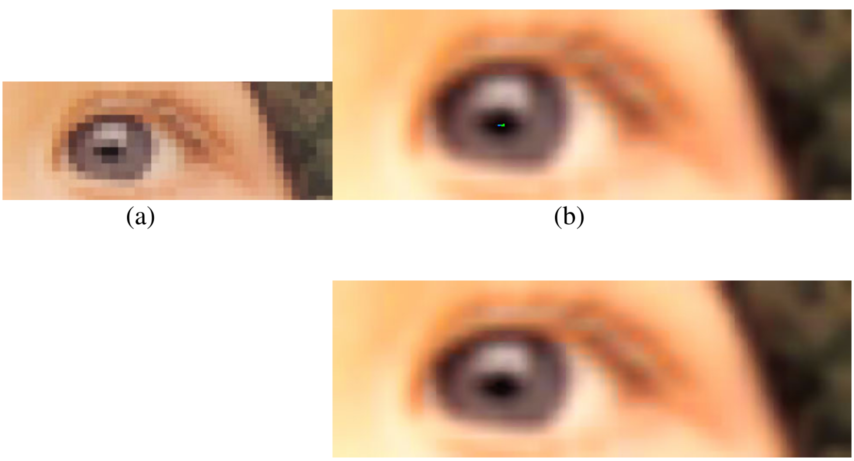
Image size \ Method	Fattal 's Method	Our Method
128×128	1.4 Sec.	0.5 Sec.
512×512	5.1 Sec.	1.4 Sec.
1024×1024	18.6 Sec.	4.8 Sec.

4.2 The Result Images of Our Method

By comparing the result images of our method with Fattal's, the result images are almost the same. The only differences was produced by the computational process of CUBLAS. Because the computation of CUBLAS uses float as its data type instead of double. Thus, there may be some problems in computational accuracy. Because the results of our method always perform the same as Fattal's, so we won't show them in our paper. In the following, we show a condition of failure, we can see the aborted result in Fig.4.

4.3 Usage in Large Images

Because of the fast and good results of our method, we can use it to process large images. But our method still needs to improve. In the future, we will use it in real-time drafting.



(a) (b) (c)

Fig. 4. (a) the input image. (b) an aborted result of our method which is produced by the computational accuracy. The failure is in the center of the child's eyes. (c) Fattal's result.

5 Conclusion

In our paper, we proposed a high-quality fast image upsampling algorithm based on CUDA. By using this method, we can generate sharp edges with reduced grid-related artifacts efficiently. We not only use edges statistics to ensure the quality of our results, but we also use CUDA to accelerate our algorithm. By using both of them, we got our high-quality fast image upsampling algorithm. We can know that the edges dependency relating certain edge features of two different resolutions. They are generically exhibited by real-world images. In addition, a special edge reliance is calculated in the preprocessing.

Acknowledgements. We'd like to thank Raanan Fattal for his help. This paper is supported by grants from the National Natural Science Foundation of China (Project No.60703070).

References

1. Shi, J.: The basis of Virtual reality and its practical algorithm. Science Press, Beijing (2002)
2. Williams, L.: Pyramidal Para metrics. Computer (1983)
3. Han, H., Zhang, C.: The research of the real-time interactive technology Clipmap in scene with huge texture, Jinan (2003)
4. Losasso, F., Hoppe, H.: Geometry clipmaps; terrain rendering using nested regular grids. ACM Transactions on Graphics 23(3) (2004)
5. Tian, D., Zheng, X.: Adaptive Large-scale Terrain texture Mapping Technology. The IADIS Computer Graphics and Visualization (2008)
6. Ratakonda, K., Ahuja, N.: POCS based adaptive image magnification. Image Processing (1998)
7. Su, Willis: Image interpolation by pixelleveldata-dependent triangulation. Computer Graphics Forum 23, 189–202 (2004)
8. Tumblin, Choudhury: Bixels: Picture samples with sharp embedded boundaries. Eurographics Symposium on Rendering (2004)
9. Fattal, R.: Upsampling via Imposed Edges Statistics. In: Proceedings of Siggraph (2007)
10. NVIDIA Corporation, http://developer.nvidia.com/

A Cubic Polynomial Model for Fisheye Camera

Haijiaing Zhu, Xiupu Yin, and Jinglin Zhou

School of Information Science & Technology,
Beijing University of Chemical Technology,
100029 Beijing, China
{zhuhj,jinglinzhou}@mail.buct.edu.cn,
2009000800@grad.buct.edu.cn

Abstract. In this paper, we present a cubic polynomial model for fisheye camera by using the lifting strategy, which point coordinates in low dimensional space is lifted to a vector in high dimensional space. In contrast to the lifting strategies reported, our lifting strategy is to let 3D point coordinates appear in higher order polynomials. This paper displays that the cubic polynomial model can effectively express the fisheye image points as the cubic polynomial of world coordinates. Thus this allows a linear algorithm to estimate the nonlinear models, and in particular offers a simple solution to estimate the nonlinear between 3D point and its corresponding fisheye image points. Experimental results with synthetic data and real fisheye images show that the fisheye camera is modeled approximately through the cubic polynomial.

Keywords: Cubic polynomial model; the lifting strategy; fisheye camera.

1 Introduction

Omnidirectional devices have been studied extensively in recent years because they can provide a wide field-of-view for photography, vision-based surveillance and virtual reality. Generally, omnidirectional cameras have two types: catadioptric cameras, which are constructed to combine a pinhole camera with a quadric mirror [1, 2], and dioptric cameras, which are composed of the pinhole camera and a fisheye lens [4-15].

There have been many works on the estimation of an omnidirectional camera model with lens distortion during more than ten years. The distortion correction models of the omnidirectional camera have several common models as follows:

The Field of View (FOV) Model [4]. This model gained an excellent distortion correction through the projection of a space line onto the camera coordinate holding a line.

The Polynomial Model. Hartley [5] expanded the radial distortion as a Taylor series and removed distortion by estimating coefficients of polynomial. Kannala [6] proposed a polynomial model between θ and r (θ is between the optical axis and the incoming ray and r is between image point and the principal point),then lens distortion correction parameters were estimated through 3D control point and corresponding 2D fisheye image point. Assuming that the image projection is described by a Taylor series expansion, Scaramuzza [7] presented a calibration technique for

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 684–693, 2011.
© Springer-Verlag Berlin Heidelberg 2011

catadioptric camera from single image and estimated the Taylor polynomial coefficients by solving a two-step least-squares linear minimization problem. Shah [8] given a polynomial transformation between 3D control points in the world coordinate and their corresponding image plane locations and calculated the polynomial coefficients using the Lagrangian estimation. Xiong [9] proposed also a polynomial model between θ and r.

The Division Model. Fitzgibbon [10] kept only the first nonlinear even term based on the polynomial model. And lens distortion parameter and epipolar geometry between multiple images were estimated simultaneously. Micusik [11] generalized Fitzgibbon's method to omnidirectional cameras with FOV than 180 degree. Suppose only point correspondences, an appropriate omnidirectional camera model was derived and estimated from epipolar geometry.

The Bicubic Model [12]. This model was described by the bicubic relation between the ideal image point and the extracted image point, and the higher order terms did not compensate for the rational distortion.

The Rational Function Model. Claus [13] proposed a rational function model and explored the mapping from image point to the incoming ray. The image point is lifted a six dimensional space, which each term is quadratic. The distortion model is then written as a matrix. For parabolic catadioptirc cameras, Geyer [14] and Barreto [15] described a distortion model through lifting image points to a four dimensional circle space. The fundamental matrix and 3D reconstruction were then estimated from two or three omnidirectional views. Sturm [16] used also lifting strategy to get model for back-projection of the incoming rays into images from affine, perspective, and para-catadioptric cameras. These models were determined to find the mapping from image points to the corresponding incidence rays. X. Ying [17] and Geyer [18] proposed a two-step projection via a quadric surface or an unit sphere, respectively. And then they tried to unify the catadioptric camera model and fisheye camera model.

Comparison with catadioptric cameras, fisheye camera model is difficult to represent by a perfect mathematical analytical function and have not been well-researched because of the severe distortions. In this paper we choose another lifting strategy to build a cubic polynomial model for fisheye camera. Our lifting strategy is to allow 3D point coordinates to appear in higher order polynomials.

This paper is in three main parts. We introduce mainly generic camera models and the lifting strategy in Section 2. The cubic polynomial model for fisheye camera is discussed in Section 3, and the mapping from 3D points to its 2D image point is described in detail. The experimental results are presented in Section 4 and Section 5 conclusion.

2 Background

There are two kinds of models for camera in the reported literatures. The first is a linear model, and the second is a nonlinear model. These models describe the mapping from 3D points to 2D points. For a linear camera, this mapping is represented by

$$\lambda \tilde{\mathbf{m}} = \mathbf{K}[\mathbf{R}, \mathbf{T}]\tilde{\mathbf{M}}_w \tag{1}$$

where $\mathbf{K}=\begin{pmatrix} f & s & u_0 \\ 0 & \alpha f & v_0 \\ 0 & 0 & 1 \end{pmatrix}$ is the intrinsic parameter matrix of camera, and f is the focal length, (u_0, v_0) is principal point, α is aspect ratio, s is skew. And $\tilde{\mathbf{M}}_w = [X_w, Y_w, Z_w, 1]^T$ is the homogeneous coordinate of 3D point in the world coordinate system, $\tilde{\mathbf{m}} = [u, v, 1]^T$ is the homogeneous coordinate of 2D point in the image plane, λ is a non-zero scale, $\mathbf{R}$ is a rotation matrix and $\mathbf{T}$ a translation vector.

For a nonlinear camera, the mapping from a 3D point to an image point includes two steps [2, 18]. A 3D point in the world coordinate is first projected linearly to a 3D point on the unit spherical image. And then the 3D point on the unit sphere is nonlinear mapped on the image plane. So, the nonlinear camera is described by

$$\lambda\tilde{\mathbf{m}} = g(\mathbf{K}\cdot[\mathbf{R},\mathbf{T}]\tilde{\mathbf{M}}_w) \tag{2}$$

where $g(\cdot)$ represents the nonlinear projection model of the camera. Recently, the rational function model has been developed to model nonlinear distortion on catadioptric and general cameras [13-16]. For parabolic catadioptric camera, authors presented an algorithm of estimation egomotion and lens geometry through lifting image points to a four-dimensional circle space [14]. And a quadratic order polynomial was used to describe the function relating a 3D ray and its corresponding image point in the omnidirectional camera [13]:

$$\boldsymbol{d}(u,v) = \boldsymbol{A}\boldsymbol{\chi}(u,v) = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \end{pmatrix} \begin{pmatrix} u^2 \\ uv \\ v^2 \\ u \\ v \\ 1 \end{pmatrix}$$

where $\mathbf{d}(u,v)$ stands for an incoming ray in R^3 corresponding to pixels (u,v) in R^2, $\mathbf{A}$ is a 3×6 matrix and $\boldsymbol{\chi}(u,v)$ is defined as the lifting point of a true image point (u,v).

This paper uses also the lifting strategy to present a nonlinear model for fisheye camera. In contrast to the lifting strategies reported, our lifting strategy is to permit 3D point coordinates to appear in higher order polynomials.

3 The Cubic Polynomial Model

The nonlinear model of the fisheye camera is derived from the lifting coordinates of 3D point in the section, and the model is the cubic polynomial model.

3.1 Mathematical Formulation

In the fisheye camera, suppose a 3D point in the world coordinate is projected on the spherical image by

$$\tilde{\mathbf{M}}_c = [\mathbf{R},\mathbf{T}]\tilde{\mathbf{M}}_w \tag{3}$$

where $\tilde{\mathbf{M}}_c = [X_c, Y_c, Z_c, 1]^T$ is the homogeneous coordinate of 3D point in the camera coordinate. And we explore the mapping of the 3D point on the camera coordinate and corresponding image point by lifting the 3D point $\tilde{\mathbf{M}}_c$ to a twenty dimensional vector

$$\begin{aligned}\boldsymbol{\Phi}(X_c,Y_c,Z_c) = [&X_c^3, Y_c^3, Z_c^3, X_c^2Y_c, X_c^2Z_c, Y_c^2X_c, Y_c^2Z_c, Z_c^2X_c, Z_c^2Y_c,\\ &X_cY_cZ_c, X_c^2, Y_c^2, Z_c^2, X_cY_c, Y_cZ_c, X_cZ_c, X_c, Y_c, Z_c, 1]^T\end{aligned} \tag{4}$$

From equations (2) and (3), our imaging model is given by

$$\lambda\tilde{\mathbf{m}} = g(\mathbf{K}\cdot\tilde{\mathbf{M}}_c) = \mathbf{P}\cdot\boldsymbol{\Phi}(X_c,Y_c,Z_c) \tag{5}$$

where $\tilde{\mathbf{m}}$ is a fisheye image point, $\mathbf{P}$ is a 3×20 matrix which is a linear combination of the distortion parameters. The nonlinear relation in equation (2) can be described by equations (4) and (5).

Let

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1,10} & \cdots & p_{1,20} \\ p_{21} & p_{22} & \cdots & p_{2,10} & \cdots & p_{2,20} \\ p_{31} & p_{32} & \cdots & p_{3,10} & \cdots & p_{3,20} \end{pmatrix}$$

From equation (5), we have

$$\lambda(u,v,1)^T = \mathbf{P}\cdot\boldsymbol{\Phi}(X_c,Y_c,Z_c) \tag{6}$$

After eliminating the scale λ from equations (4), (5) and (6), we get

$$\begin{cases} u = \dfrac{P_1(X_c,Y_c,Z_c)}{P_3(X_c,Y_c,Z_c)} \\ v = \dfrac{P_2(X_c,Y_c,Z_c)}{P_3(X_c,Y_c,Z_c)} \end{cases} \tag{7}$$

where

$$\begin{aligned}P_1(X_c,Y_c,Z_c) = {}& p_{11}X_c^3 + p_{12}Y_c^3 + p_{13}Z_c^3 + p_{14}X_c^2Y_c + p_{15}X_c^2Z_c + p_{16}Y_c^2X_c + p_{17}Y_c^2Z_c \\ &+ p_{18}Z_c^2X_c + p_{19}Z_c^2Y_c + p_{1,10}X_cY_cZ_c + p_{1,11}X_c^2 + p_{1,12}Y_c^2 + p_{1,13}Z_c^2 \\ &+ p_{1,14}X_cY_c + p_{1,15}Y_cZ_c + p_{1,16}X_cZ_c + p_{1,17}X_c + p_{1,18}Y_c + p_{1,19}Z_c + p_{1,20}\end{aligned}$$

$$\begin{aligned}P_2(X_c,Y_c,Z_c) = {}& p_{21}X_c^3 + p_{22}Y_c^3 + p_{23}Z_c^3 + p_{24}X_c^2Y_c + p_{25}X_c^2Z_c + p_{26}Y_c^2X_c + p_{27}Y_c^2Z_c \\ &+ p_{28}Z_c^2X_c + p_{29}Z_c^2Y_c + p_{2,10}X_cY_cZ_c + p_{2,11}X_c^2 + p_{2,12}Y_c^2 + p_{2,13}Z_c^2 \\ &+ p_{2,14}X_cY_c + p_{2,15}Y_cZ_c + p_{2,16}X_cZ_c + p_{2,17}X_c + p_{2,18}Y_c + p_{2,19}Z_c + p_{2,20}\end{aligned}$$

$$P_3(X_c,Y_c,Z_c) = p_{31}X_c^3 + p_{32}Y_c^3 + p_{33}Z_c^3 + p_{34}X_c^2Y_c + p_{35}X_c^2Z_c + p_{36}Y_c^2X_c + p_{37}Y_c^2Z_c \\ + p_{38}Z_c^2X_c + p_{39}Z_c^2Y_c + p_{3,10}X_cY_cZ_c + p_{3,11}X_c^2 + p_{3,12}Y_c^2 + p_{3,13}Z_c^2 \\ + p_{3,14}X_cY_c + p_{3,15}Y_cZ_c + p_{3,16}X_cZ_c + p_{3,17}X_c + p_{3,18}Y_c + p_{3,19}Z_c + p_{3,20}$$

3.2 Estimation of Model Parameters

Given a set of 3D space point to image point correspondence $\tilde{\mathbf{M}}_c \leftrightarrow \tilde{\mathbf{m}}$, the distortion parameters $\mathbf{P}$ can be computed, namely the coefficients of polynomials $P_1(X_c,Y_c,Z_c)$, $P_2(X_c,Y_c,Z_c)$ and $P_3(X_c,Y_c,Z_c)$. From equation (7), we get

$$\begin{cases} u \cdot P_3(X_c,Y_c,Z_c) - P_1(X_c,Y_c,Z_c) = 0 \\ v \cdot P_3(X_c,Y_c,Z_c) - P_2(X_c,Y_c,Z_c) = 0 \end{cases} \tag{8}$$

Two constraints on elements of $\mathbf{P}$ matrix are firstly obtained from equation (8). Given n correspondences between world points and image points, we have $2n$ constraints. By stacking all constraints, a system of linear equation is then rewrite as

$$\begin{pmatrix} X_{c1}^3, Y_{c1}^3, \cdots, 1, 0, 0, \cdots, 0, -u_1X_{c1}^3, -u_1Y_{c1}^3, \cdots, -u_1 \\ 0, 0, \cdots, 0, X_{c1}^3, Y_{c1}^3, \cdots, 1, -v_1X_{c1}^3, -v_1Y_{c1}^3, \cdots, -v_1 \\ \vdots, \vdots, \cdots, \vdots, \vdots, \vdots, \cdots, \vdots, \vdots, \vdots, \cdots, \vdots \\ X_{cn}^3, Y_{cn}^3, \cdots, 1, 0, 0, \cdots, 0, -u_nX_{cn}^3, -u_nY_{cn}^3, \cdots, -u_n \\ 0, 0, \cdots, 0, X_{cn}^3, Y_{cn}^3, \cdots, 1, -v_nX_{cn}^3, -v_nY_{cn}^3, \cdots, -v_n \end{pmatrix} \cdot \mathbf{C} = 0 \tag{9}$$

where $\mathbf{C} = (p_{11}, p_{12}, \cdots, p_{1,20}, p_{21}, p_{22}, \cdots, p_{2,20}, p_{31}, p_{32}, \cdots, p_{3,20})^T$ is a 60-dimensional vector. There are a lot of 60 unknown parameters in equation (9), and 30 points correspondences are at least required to solve this equation. These coefficients can be solved from the linear least-squares minimization [5].

Similarly, the matrix between 3D points and image points can be computed by equation (5). These camera parameters are refined by minimizing an objective function that is a sum of squared distances between the measured and re-projected point using the Levenberg-Marquardt algorithm. The objective function is built by the following step.

1. Compute the 3×20 matrix $\mathbf{P}$ by a set of 3D space point to image point correspondence $\tilde{\mathbf{M}}_c \leftrightarrow \tilde{\mathbf{m}}$ and equation (8).
2. Estimate the reprojected image point $\tilde{\mathbf{m}}'$ using equation (7).
3. Suppose there are N image points in the fisheye image, then the objective function is

$$\tau = \sum_{i=1}^{N} d(\tilde{\mathbf{m}}_i \tilde{\mathbf{m}}_i')^2 \tag{10}$$

3.3 Summary

Given the homogeneous coordinates of 3D point in the camera coordinate, and algorithm for computing the cubic polynomial model is as follows:

1. Take a few fisheye images of the 3D calibration object.
2. Detect the corner points in the fisheye images.
3. Build the lifting 3D coordinates from equation (4).
4. Calculate the 3×20 matrix $\mathbf{P}$, and get an initial estimation of the coefficients of the cubic polynomial model.
5. Refine these coefficients by minimizing the objective function using the Levenberg-Marquardt algorithm.

4 Experimental Results

In the experiment, we choose the equidistance projection as the projection of fisheye camera. Synthetic experiments are firstly presented, and experiments with real fisheye image are given.

4.1 Experiments with Synthetic Data

In synthetic experiment, the approximate projection model of the fisheye camera is chosen as:

$$r \approx f\theta + k_1\theta^2 + k_2\theta^3 + k_3\theta^4 + k_4\theta^5$$

where f is the focal length, θ is the angel between the incoming ray and the optical axis and r is the distance from the image point to the principal point.

The intrinsic parameters of the simulated fisheye camera is

$$f = 150.0\,, \alpha = 1.0\,, u_0 = 320\,, v_0 = 240\,, k_1 = 15.0\,, k_2 = -20.0\,, k_3 = 2.0\,, k_4 = -10.0\,.$$

And the extrinsic parameters of two simulated fisheye images are

$$R_1 = \begin{pmatrix} 0.9980 & 0.0092 & 0.0618 \\ -0.0072 & 0.9995 & -0.0312 \\ -0.0621 & 0.0307 & 0.9976 \end{pmatrix}, T_1 = \begin{pmatrix} -232.3595 \\ -182.8318 \\ -492.5258 \end{pmatrix}$$

$$R_2 = \begin{pmatrix} 0.9982 & 0.0415 & 0.0442 \\ -0.0388 & 0.9974 & -0.0602 \\ -0.0466 & 0.0584 & 0.9972 \end{pmatrix}, T_2 = \begin{pmatrix} -234.4509 \\ -155.5824 \\ -600.6887 \end{pmatrix}$$

Simulated 3D space points are shown in Fig. 1, and 64 place points on XOY, YOZ and XOZ plane are taken, respectively. These 3D points are linearly projected to the unit sphere through the extrinsic parameters, and these points on the unit sphere are mapped on the fisheye image by using the intrinsic parameters of synthetic fisheye camera. Two simulated images are shown in Fig. 2 and their sizes are 1000×1000 pixels.

We take first higher order coordinate of 3D points (shown as in Fig. 1) using our lifting strategy, and the 3×20 matrix $\mathbf{P}$ is estimated by a set of image points in Fig. 2 and the lifting higher coordinates of 3D points. After estimating the $\boldsymbol{P}$, the 3D points are re-projected onto the image plane, and all image points are marked by black "+" in Fig. 3 (a) and (b). We can see that the lines in the synthetic image are almost become straight lines. These experimental results show that this model is feasible and reliable.

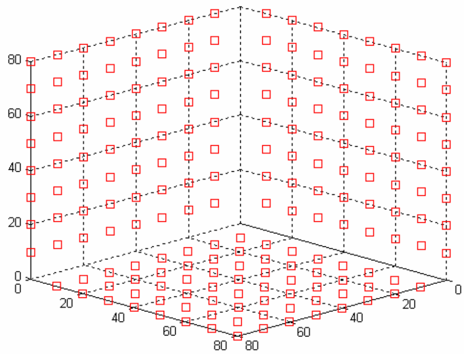
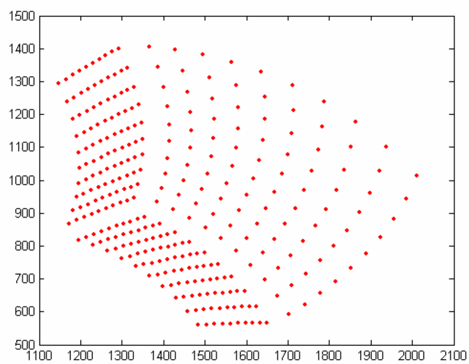
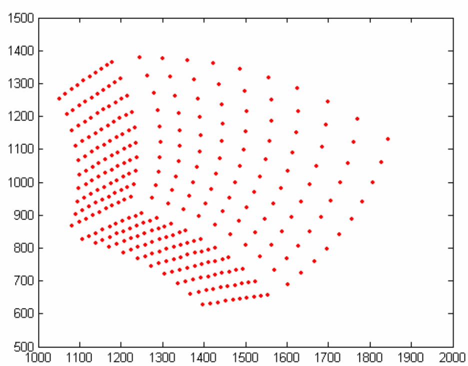


Fig. 1. The synthetic 3D space points

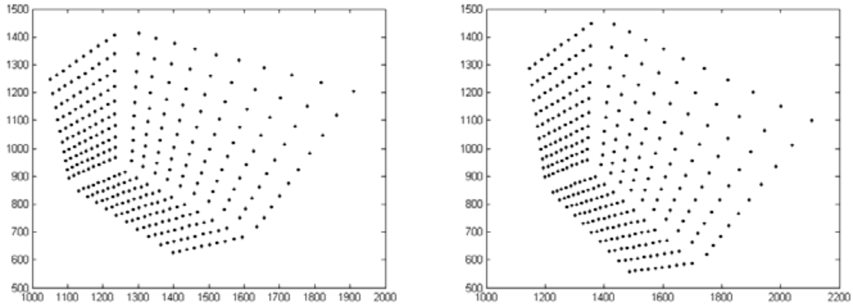


(a)



(b)

Fig. 2. Two simulated fisheye images



(a) (b)

Fig. 3. Corrected image of the synthetic fisheye image in Fig. 2

4.2 Experiments with Real Fisheye Image

Multiple fisheye images were taken to use a fixed camera with fisheye lens in front of a 3D calibration object. One of these images, with a size of 640×480 pixels, is shown in Fig. 4. In this experiment, 3D coordinates of corner points in the 3D calibration object are given and fisheye image points are detected using the OpenCV function. Then the 3×20 matrix $\mathbf{P}$ is estimated by using algorithm in Section 3 from a set of 3D points and image 2D point correspondences.

After estimated the projective matrix, we can calculate the projected point in fisheye image using equation (7). Then the real fisheye image in Fig. 4 is corrected and it is shown as Fig. 5. We can see that those heavily distorted lines around the fisheye image have become almost straight lines. These experimental results show that the cubic polynomial function can be modeled the fisheye camera.

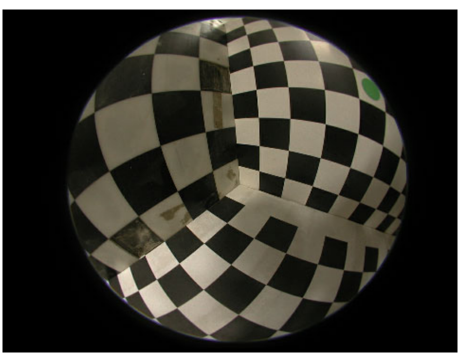


Fig. 4. One of multiple real fisheye images

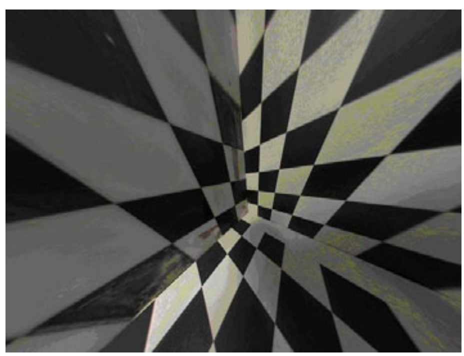


Fig. 5. Corrected fisheye image

In addition, some fisheye images of lab indoor scene are taken, and their sizes are 640×480 pixels. One of these fisheye images is shown in Fig. 6(a), and points marked by red are used to calculate the parameter matrix. After estimating the projective matrix, the real fisheye image is corrected and shown in Fig. 6(b). From this result, we know that these heavy distorted lines on the ceiling, floor and door are almost become straight lines.

(a) (b)

Fig. 6. (a) One fisheye image of lab indoor scene, and (b) corrected fisheye image

5 Conclusion

We have extended the lifting strategies from the 2D image points to the 3D points and described a cubic polynomial model for fisheye camera in this paper. Given a set of 3D point to 2D image point correspondences, our lifting strategy was implemented to obtain higher coordinates on 3D points. And a linear combination of the distortion parameters of fisheye camera was estimated by our method. Then, experiments with synthetic and real fisheye image were implemented, and the results have shown the cubic polynomial model for fisheye camera is reliable.

Acknowledgements. This work was supported by the National Natural Science Foundation of China under grant No.60875023, The Fundamental Research Funds for the Central Universities (ZZ1013) and The Project-sponsored by SRF for ROCS, SEM. We thank Dr. Dong Qiulei in NLPR, Institute of Automation Chinese Academy Sciences, who gave us some real fisheye images.

References

1. Nayar, S.: Catadioptric Omnidirectional Camera. In: IEEE Conference Computer Vision and Pattern Recognition, pp. 482–488. IEEE Computer Society Press, Puerto Rico (1997)
2. Geyer, C., Daniilidis, K.: Catadioptirc Projective Geometry. International Journal of Computer Vision 45(3), 223–243 (2001)
3. Robert, W.: Physical Optics. The Macmillan Company, New York (1911)
4. Devernay, F., Faugeras, O.D.: Straight lines have to be straight. Machine Vision and applications 13, 14–24 (2001)
5. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision, 2nd edn. Cambridge Press, Cambridge (2003)
6. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventenal, wide-eye, and fish-eye lenses. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(8), 1335–1340 (2006)

7. Scaramuzza, D., Martinelli, A., Siegwart, R.: A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. In: Proceedings of IEEE International Conference of Vision Systems. IEEE Computer Society Press, New York (2006)
8. Shishir Shah, J.K., Aggarwal, A.: A Simple Calibration Procedure for Fish-Eye (High-Distortion) Lens Camera. In: Proceedings of International Conference on Robotics and Automation, San Diego, CA, USA, May 1994, vol. 4, pp. 3422–3427. IEEE Computer Society Press, Los Alamitos (1994)
9. Xiong, Y., Turkowski, K.: Creating Image-Based VR Using a Self-Calibrating Fisheye Lens. In: IEEE Conference Computer Vision and Pattern Recognition, pp. 238–243. IEEE Computer Society Press, Puerto Rico (1997)
10. Fitzgibbon, A.W.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: Proceedings of Computer Vision and Pattern Recognition, pp. 125–132. IEEE Computer Society Press, Kauai (2001)
11. Micusik, B., Pajdla, T.: Estimation of omnidirectional camera model from epipolar geometry. In: Proceedings of Computer Vision and Pattern Recognition, pp. 485–490. IEEE Computer Society Press, Madison (2003)
12. Kilpel'a, E.: Compensation of systematic errors of image and model coordinates. In: International Archives of Photogrammetry, pp. 407–427 (1980)
13. Claus, D., Fitzgibbon, A.W.: A rational function lens distortion model for general cameras. In: Proceedings of Computer Vision and Pattern Recognition, pp. 213–219. IEEE Computer Society Press, San Diego (2005)
14. Geyer, C., Daniilidis, K.: Structure and motion from uncalibrated catadioptric views. In: Proceedings of Computer Vision and Pattern Recognition, pp. 279–286. IEEE Computer Society Press, Kauai (2001)
15. Barreto, J.P.: Fundamental matrix for cameras with radial distortion. In: IEEE International Conference on Computer Vision, pp. 625–632. IEEE Computer Society Press, San Diego (2005)
16. Sturm, P., Ramalingam, S.: A generic concept for camera calibration. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3022, pp. 1–13. Springer, Heidelberg (2004)
17. Ying, X., Hu, Z.: Can We Consider Central Catadioptric Cameras and Fisheye Cameras within a Unified Imaging Model. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 442–455. Springer, Heidelberg (2004)
18. Geyer, C., Daniilidis, K.: A Unifying Theory for Central Panoramic Systems and Practical Implications. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 445–461. Springer, Heidelberg (2000)

Author Index

Abdullah, Natrah I-17
Abe, Kiyohiko II-176
Abe, Shuya III-240
Abou Khaled, Omar II-185, II-222, III-165
Abraham, Chon IV-141
Abril-Jiménez, Patricia III-129
Abusham, Eimad E.A. II-169
Adams, Candice IV-3
Adikari, Sisira I-25
Ahn, Hyeong-Joon III-386
Ahn, Jong-gil II-304
Ai, Renlong II-387
Aihara, Kenro III-271
Akahane, Katsuhito II-3
Akaho, Kengo III-357
Alcañiz, Mariano I-423
Alnizami, Hanan II-533
Altaboli, Ahamed I-35
Alvarez, Haydee III-129
Amraji, Nitish I-539
Andrade, Vladimir I-54
Andreoni, Giuseppe I-406
Andrich, Rico II-396
Angelini, Leonardo II-185
Antonya, Csaba II-204
Aoki, Ryosuke II-194
Aoyagi, Saizo III-465
Aoyama, Hifumi III-84
Aoyama, Hisae I-341
Arafat, Sachi III-367
Arredondo Waldmeyer, María Teresa III-129, IV-219
Asan, Onur II-13
Asano, David II-133
Ash, Jordan III-137
Auernheimer, Brent IV-394

Babes, Monica III-137
Bachman, Mark I-107, IV-151
Balasubramanian, Venkatesh II-446
Baños, Rosa I-423
Baragaño Galán, José R. III-129
Baranauskas, M. Cecília C. I-72
Barbuceanu, Florin II-204
Barnes, Julie I-432
Barrett, Jonathan IV-580
Barroso, João III-293
Bartoschek, Thomas II-71
Bashir, Housam K. II-169
Beheshti, Jamshid IV-541, IV-590
Bengler, Klaus II-23, II-125, III-376
Benoit, David Michel II-537
Bernstein, Stefan I-618
Beznosyk, Anastasiia IV-668
Bilal, Dania IV-549
Birinci, Murat I-547
Biswas, Gautam IV-580
Biswas, Pradipta IV-425
Bizik, Kai I-165
Black, Alan II-358
Blackwell, Nicole II-547
Blattner, Andreas II-125
Blažica, Bojan III-519
Block, Micha III-474
Blumberg, Malte III-38
Böck, Ronald III-603
Boegel, Bastian II-537
Bomsdorf, Birgit I-155
Bonarini, Andrea IV-649
Botella, Cristina I-423
Boulabiar, Mohamed-Ikbel II-214
Bowden, Chris IV-443
Braun, Andreas III-147, III-205
Breiner, Kai I-165, I-299
Breyer, Matthias IV-520
Bricault, Ivan IV-202
Britton, Douglas IV-167
Brooks, Laurence IV-456
Bubb, Heiner III-376
Bucki, Marek IV-202
Buckwalter, J. Galen II-474
Burger, Thomas II-214
Burkhardt, Dirk IV-520
Byer, David II-325

Camez, Edgar III-411
Campbell, John I-25

Campos, Miguel III-155
Campos, Pedro III-155
Caon, Maurizio II-185
Carrino, Francesco II-222
Carrino, Stefano II-185, III-165
Casillas, Raúl III-3
Castilla, Diana I-423
Cearreta, Idoia III-525
Cerrato, Marcus Vinicius I-471
Chaim, Marcos Lordello I-471
Chalon, René II-500
Chamaret, Damien II-158
Chan, Foong-Yeen Donny III-329
Chance, Eric II-474
Chang, Huang-Ming III-175, III-559
Chang, Liang I-555, I-670
Chang, Teng-Wen I-82, II-53, II-409, III-250
Chao, Dingding III-280
Charfi, Selem I-175
Charissis, Vassilis III-367
Chattaraman, Veena II-533
Chen, Chun-Lien IV-570
Chen, Jai-Jung I-82
Chen, Jie I-677
Chen, Shang-Liang IV-570
Chen, Shao-Nung IV-86
Chen, Sheng-Han II-409
Chen, Sherry Y. IV-27
Chen, Ting-Han III-185
Chen, Wei-Chia III-21
Chen, Yihsiu III-13
Chen, Yun-Yao IV-570
Cheng, Yi-Ju IV-131
Cheng, Yun-Maw III-21
Chiang, Chen-Wei I-45, III-30
Chiu, Kuo Chung II-491
Cho, Hyunchul II-119
Cho, Young Suk IV-192
Choe, Jaeho III-401
Choi, Ji-Hye I-637, I-653
Choi, Jongwoo III-452
Choi, Seung-Hwan II-231
Chou, Chin-Mei I-397
Chu, Chi Nung III-288, IV-35, IV-435
Chu, Gene III-288
Chu, Keng-Yu II-248
Chu, Shaowei II-238
Chuang, Chih-Fei IV-439
Cobb, Sue III-483
Cockton, Gilbert III-490
Coelho, Paulo III-293
Cohen, Gal III-137
Coninx, Karin IV-668
Cooper, Eric W. III-535
Coppin, Gilles II-214
Cortés, José Luis IV-255
Costa, Fiammetta IV-649
Costa, Paulo III-293
Couvreur, Christophe III-418
Cremers, Anita II-575

Dadgari, Darius II-331
Damböck, Daniel III-376
Dandekar, Heramb III-411
Datta, Sambit I-185
David, Bertrand II-500, III-301
D'Cruz, Mirabelle III-483
de Barrionuevo, Arturo Díaz IV-219
de Beer, Lorraine III-500
Dees, Walter III-195
de Moraes Rodrigues, Roberto Leite I-471
Deng, Qingqiong I-670
Deng, Xiaoming I-555
Depradine, Colin II-325
Diebold, Philipp I-165
Dittmar, Anke I-194
Djamasbi, Soussan I-331, IV-245
Do, Ellen Yi-Luen II-547, IV-192
Donker, Hilko III-38
Druin, Allison IV-559
Drury, Colin G. I-397, II-275
Duan, Fuqing I-660, I-670
Duguleana, Mihai II-204

Edlinger, Günter II-417
Edlin-White, Rob III-483
Ei, Hitomi IV-413
Ekman, Inger I-351
Ellis, Rebekah IV-549
EL-Qirem, Fuad III-490
Engel, Jürgen I-204
Esteves, Marcela I-54
Everard, Andrea IV-58, IV-235
Ezzedine, Houcine I-175

Fahn, Chin-Shyurng II-248
Fails, Jerry Alan IV-559
Fan, Kuo-Kuang III-259

Fang, Xiaowen IV-659
Feki, Mohamed Ali I-229
Fennell, Antoinette IV-443
Fernandes, Hugo III-293
Fernandez-Llatas, Carlos I-214
Fernández, Cristóbal IV-255
Ferran, Nuria III-510
Feuerstack, Sebastian I-221
Fiddian, Thomas IV-443
Fischer, Holger II-100
Floyde, Anne III-483
Folmer, Jens IV-297
Forbrig, Peter I-194, I-204, I-248, I-319, IV-510
Forutanpour, Babak III-48
Fox, Michael A. IV-307
Friedman, Nizan IV-151
Friesen, Rafael I-121, III-593
Frommer, Jörg III-593
Fuji, Masahiko IV-21
Furukawa, Hiroyuki IV-159
Furukawa, Masashi IV-354
Furukawa, Yutaroh I-588
Furuta, Kazuo III-280
Furuyama, Nobuhiro II-259

Gabbouj, Moncef I-547
Gacimartín, Carlos I-443
Gagliardi, Chiara I-406
Gandy, Maribeth IV-167
Gao, Fei III-59
Gao, Qin I-450, III-59
Gao, Song-feng IV-688
Garay-Vitoria, Nestor III-525
Garbarino, Maurizio IV-649
Garreta-Domingo, Muriel I-136, III-510
George, Sébastien IV-12
Germanakos, Panagiotis I-388
Ghinea, Gheorghita II-427
Gilbert, Juan II-533
Gill, Sukhpreet IV-394
Glodek, Michael III-603
Godon, Marc I-229
Gonçalves, Júnia I-63
Gong, Yang IV-182
Gower, Michael I-248
Grau, Stefan I-155
Grønli, Tor-Morten II-427
Grundlehner, Bernard II-481
Guger, Christoph II-417
Guha, Mona Leigh IV-559
Guldogan, Esin I-547
Guo, Ping I-562

Haase, Matthias III-593
Hadjileontiadis, Leontios III-293
Hamada, Takeo II-3
Hamberger, Werner II-125
Han, Ji-Hyeong II-231
Han, Tack-don I-637, I-653
Han, Yanjun I-555
Hanada, Ryoko II-259
Hansen, Jarle II-427
Harbert, Simeon IV-167
Harley, Linda IV-167
Harrelson, Heather III-411
Hashimoto, Yasuhiro IV-328
Haslbeck, Andreas II-23
Hatano, Akira I-628
Hayakawa, Eiichi IV-21
Hayashi, Yugo II-513, IV-384
He, Ning I-571
Heidemann, Gunther I-618
Heikkinen, Kari I-359
Hermann, Thomas II-110
Hernández, José Alberto I-443
Hess, Steffen I-238
Hijikata, Yoshinori IV-318
Hippler, Rachelle Kristof II-33
Hiramatsu, Mizuko III-84
Hobbs, Michael I-185
Höger, Rainer III-577
Holzner, Clemens II-417
Honda, Daisuke III-230
Hong, Chao-Fu IV-77
Hong, Kwang-Seok II-464, III-569
Hope, Tom IV-328
Horii, Satoshi III-84
Hornung, Heiko I-72
Howley, Iris II-341
Hrabal, David III-603
Hsiao, Chih-Pin II-43
Hsu, Chia-Ling IV-77
Hsu, Chih-Hsiang IV-453
Hsu, Yung-Chi IV-27
Hu, Xiaoyan I-580
Huang, Ho-Chuan IV-570
Huang, Hsiu-Shuang IV-570
Huang, Nancy IV-439
Huang, Yu-Chun Annester I-82

Huang, Zhao IV-456
Hudasch, Martin I-155
Hung, Tien-Hsin I-82
Huseyinov, Ilham N. IV-39
Hwang, Bong-Ha III-391
Hwang, Sheue-Ling III-435, IV-439
Hwang, Wonil III-386
Hwang, Yu Ting IV-35
Hwangbo, Hwan III-391
Hymel, Alicia M. IV-580
Hyrkkänen, Ursula III-69

Ichinomiya, Takako II-259
Ichiyanagi, Yuki III-535
Igata, Nobuyuki III-240
Ihara, Masayuki II-194
Ike, Shinichi II-436
Imtiaz Ahmad, Muneeb III-543
Ingold, Rolf II-222, III-165
Inoue, Masashi II-259
Inoue, Satoru I-341
Irino, Toshio II-259, IV-607
Ishida, Hirotsugu I-588, I-628, IV-174, IV-210
Ishida, Toru II-565
Ishidou, Takashi IV-413
Ishii, Hirotake III-465, IV-403
Ishii, Naohiro I-608
Islam, Md. Zahidul II-294
Ito, Kyoko IV-48
Ito, Yuki IV-48
Itoh, Kazunori IV-97
Izuhara, Ritsuko IV-237
Izumiya, Akira IV-226

Jacob, Robert J.K. II-593
Jaffee, Samuel D. I-432, II-33
Jalal, Sameen III-137
Jang, Bong-gyu III-76
Jang, Seung-Cheol IV-360
Javahery, Homa I-248
Jayaraman, Srinivasan II-446
Jeon, Myounghoon II-523
Ji, Yong Gu III-391
Jiang, Heng II-53
Jin, Beom Suk III-391
Jin, Byungki IV-257
Jin, Hong Zhe IV-403
Jin, Zhaoxia Janet III-329
Jog, Jayraj II-148
Johnson, Brian R. II-43
Johnson, Nathan IV-336
Johnson, Steve II-62, IV-68
Jokisch, Markus II-71
Jones, Brian M. IV-58, IV-235
Joo, Jaekoo I-368
Jorge, Joaquim III-155
Jung, Eui S. III-401, III-426
Jung, Hee-Seok III-386
Jung, Helge II-100
Junglas, Iris IV-141

Kaelber, Claus I-258
Kagami, Shingo II-194
Kallenbach, Jan IV-466
Kallinen, Kari I-351, IV-466
Kalwar, Santosh Kumar I-359
Kamieth, Felix III-147, III-205
Kaneko, Ryuhei IV-346
Kang, Sunghyun R. IV-285
Kanno, Taro III-280
Karatsu, Yutaka II-194
Karpov, Alexey II-454
Karthik, K. III-552
Kato, Hirokazu III-357
Kawahara, Hideki IV-607
Kawai, Katsuya III-357
Kawakami, Takashi IV-617
Kawano, Show IV-403
Kayama, Mizue IV-97
Keller, Christine I-288
Khairat, Saif IV-182
Kienle, Martin III-376
Kim, Dong-Kyu II-464
Kim, Gerard J. II-304, III-76
Kim, Hee-Cheol IV-476
Kim, Ho-Won IV-476
Kim, Hyungsin IV-192
Kim, Jaewhan IV-360
Kim, Ji Yeon III-391
Kim, Jong-Hwan II-231
Kim, Jung Hyup I-378
Kim, Kyong-ho III-452
Kim, Laehyun II-119
Kim, Namhun I-368
Kim, Nam-Hyo III-386
Kim, Sang-Hwan III-411
Kim, Seong M. III-401
Kim, Taeil III-426
Kimioka, Gimpei II-81

Kimura, Hiroshi IV-354
Kimura, Masaomi I-588, I-628, IV-174, IV-210, IV-226
Kipyatkova, Irina II-454
Kirisci, Pierre T. IV-443
Kita, Yusuke III-213
Klein, Brandi A. II-33
Kleindienst, Jan III-418
Klenner-Moore, Jayne IV-61
Klompmaker, Florian II-90, II-100
Klopfer, Dale S. I-432, II-33
Kobayashi, Minoru II-194
Koga, Hiroyuki I-461
Kogure, Yuri IV-413
Kojima, Kazuaki II-613, IV-384
Kojo, Akira I-588
Kolski, Christophe I-175, III-301
Komischke, Tobias I-92
Konuma, Mizuki IV-413
Koshi, Shoko II-314
Koshiba, Hitoshi III-271
Krahmer, Emiel II-575, III-543
Krause, Michael II-23
Kryssanov, Victor V. II-513, III-535
Kuijper, Arjan I-480, IV-520
Kuramoto, Itaru II-351
Kurihara, Kazutaka II-557
Kurihara, Lisa IV-607
Kweon, In-So II-294
Kwon, Hyeong-Joon III-452
Kwon, Hyeong-Oh III-569
Kwon, Wi-Suk II-533

Labský, Martin III-418
Lang, Helmut II-537
Langdon, Patrick IV-425
Lange, Christian II-125
Lange, Julia III-593
Large, Andrew IV-541, IV-590
Larrabeiti, David I-443
Lawo, Michael IV-443
Le Bellego, Gaël IV-202
Lee, Chao-Lung III-21
Lee, Chil-Woo II-268, II-294
Lee, Hyunglae IV-484
Lee, Jun-Sung II-268, II-294
Lee, Myeong-Hee IV-328
Lee, Myonghee III-76
Lee, Seung Jun IV-360
Lee, Ying-Lien IV-453
Lekkas, Zacharias I-388
Leou, Yea-Mei IV-570
Leventhal, Laura Marie I-432, II-33
Levin, Daniel T. IV-580
Lewthwaite, Sarah III-483
Li, Yibai IV-336
Li, Yueqing II-62, IV-68
Li, Zhanhuai III-220
Liang, Chao I-660
Liang, Rung-Huei III-175, III-559
Liang, Sheau-Farn Max I-98
Lichtenberg, Sam III-137
Lim, Youngjae III-426
Limbrecht, Kerstin III-603
Lin, Brian Tsang-Wei III-435
Lin, Chien-Tsen I-98
Lin, Hsin-Hou III-185
Lin, Jui-Feng I-397, II-275
Lin, Mu-Hua IV-77
Lin, Sheng Ta II-491
Lin, Ya-Li III-311
Lin, Yi-Chu III-559
Lin, Yingzi I-35
Lin, Yi-Quan I-397
Lin, Yu-De I-397
Lin, Yuh-Chang IV-77
Link, Jasmin III-474
Littman, Michael III-137
Liu, Cha-Lin II-53
Liu, Cheng-Li IV-490
Liu, Deyun IV-476
Liu, Jane W.-S. III-107
Liu, Jie IV-368
Liu, Xianging II-140
Liu, Yen Ping II-491
Loiacono, Eleanor IV-245
López, Marta I-136
Loveless, Steve IV-394
Lu, Cindy IV-375
Lu, Ke I-571
Lu, Szu-Chia II-547
Lu, Yin III-220
Luh, Ding-Bang IV-86
Luo, Ruiyi I-598
Luo, Wei Wen II-491
Luyten, Kris IV-668

Ma, Jin I-562
Macek, Tomáš III-418
Maciejewski, Anthony A. I-450

Mackenzie, Lewis III-367
Maeda, Atsuhiko II-194
Maehigashi, Akihiro IV-384
Maekawa, Yasuko III-84
Magennis, Mark IV-443
Maier, Andreas I-238
Majima, Yukie III-84
Mallem, Malik II-140
Mangas, Juan-Antonio III-510
Marivate, Vukosi III-137
Märtin, Christian I-204, I-258
Maruyama, Minoru II-133
Massaki, Hiroyasu II-259
Matos, Gilberto I-142
Matsui, Tatsunori II-613
Matsunobe, Takuo II-378
Matsuuchi, Naohisa II-436
Matteucci, Matteo IV-649
Mattheij, Ruud III-500
Mayer, Jürgen II-23
Mazalek, Ali II-148
McCoy, Scott IV-58, IV-235, IV-255
McDonald, Craig I-25
Md Noor, Nor Laila I-17
Meixner, Gerrit I-165, I-299
Mendenhall, Sam II-148
Merlo, Mark I-107
Metze, Florian II-358
Miao, Kejian III-220
Milanova, Mariofanna I-539
Milde, Jan-Torsten I-155
Min, Jin-Hong III-569
Minakuchi, Mitsuru II-351
Minker, Wolfgang I-527, II-537
Mitchell, Stephen J. IV-394
Mitsuishi, Takashi IV-500
Miwa, Kazuhisa II-513, IV-384
Miyabe, Mai II-368
Miyagi, Kazune IV-403
Miyamori, Shoko IV-607
Miyao, Hidetoshi II-133
Miyazaki, Tsuyoshi I-608
Miyoshi, Yasuo IV-93
Mizukawa, Makoto III-586
Mladenić, Dunja III-519
Mochizuki, Toshio II-557
Mocholí, Juan Bautista I-214
Moehrmann, Julia I-618
Mohamad, Yehya IV-443
Mohd. Ali, Borhanuddin II-285
Montague, Enid II-13
Moon, Young Joo III-391
Mor, Enric III-510
Moragrega, Inés I-423
Morán, Alberto L. III-3
Morandini, Marcelo I-471
Morie, Jacquelyn Ford II-474
Morita, Junya IV-384
Mort, Greg III-321
Mortimer, Bruce III-321
Mosch, Christian II-537
Mourlas, Constantinos I-388
Moussa, Faouzi I-175, III-92
Mu, Lin I-539
Mugellini, Elena II-185, II-222, III-165
Müller, Tobias II-90
Muranaka, Noriaki IV-121
Muraoka, Hiroki IV-226
Murata, Kazuyoshi III-101
Murata, Yuichi II-557

Nabeta, Keita I-588, I-628, IV-174, IV-210
Nagai, Takashi IV-97
Nagata, Kazunobu III-101
Naka, Toshiya II-565
Nakada, Toru II-481
Nakagawa, Takashi III-357
Nakagawa, Yuuki IV-21
Nakajima, Yukari III-84
Nakamura, Yuki III-444
Nakamura, Yumiko III-84
Nakashima, Toyoshiro I-608
Nakata, Keiichi I-341
Nakatani, Yoshio III-213, III-444, IV-346
Nam, Chang Soo II-62, IV-68
Naranjo, Juan Carlos I-214
Nauts, Pim II-575
Nazemi, Kawa I-480, IV-520
Nebe, Karsten I-114, II-90, II-100
Nenonen, Suvi III-69
Nesset, Valerie IV-599
Ng, Chee Kyun II-285
Ng, Wei Lun II-285
Nishida, Shogo III-230, III-357, IV-48, IV-318
Nishino, Yosuke IV-21
Nisimura, Ryuichi IV-607
Noordin, Nor Kamariah II-285

Ochiai, Hideharu IV-21
O' Connor, Joshue IV-443
Octavia, Johanna Renny IV-668
Oehl, Michael III-577
Ogawa, Hitoshi II-513, III-535
Oh, Chi-Min II-268, II-294
Ohi, Shoichi II-176
Ohkawa, Yuichi IV-500
Ohkura, Michiko I-588, I-628, IV-174, IV-210, IV-226, IV-413
Ohyama, Minoru II-176
Oka, Mizuki IV-328
Okamoto, Kentaro II-585
Okamura, Tomoaki III-465
Okazaki, Tetsuo IV-617
Okuya, Ryo IV-210
Oliveira, Kathia III-301
Olivier, Hannes IV-678
Omernick, Mark II-13
Oobayashi, Takaaki IV-93
Otmane, Samir II-140
Ottaviano, Manuel IV-219
Otto, Mirko I-121, III-593
Ozaki, Shun II-378

Paelke, Volker I-114
Páez, José Manuel IV-219
Pantförder, Dorothea IV-297
Papanastasiou, Stylianos III-367
Park, Gie-seo II-304
Park, Hyesun III-452
Park, Jaekyu III-401
Park, James I-637
Park, Ji-Hyung IV-484
Park, Sehyung II-119
Park, Sungjoon III-426
Park, Wanjoo II-119
Patki, Shrishail II-481
Pavlov, Oleg IV-245
Pecot, Katrina II-23
Peer, Dain II-13
Peinado, Ignacio IV-219
Peissner, Matthias I-268, I-498
Penders, Julien II-481
Penstein Rosé, Carolyn II-341
Perego, Paolo I-406
Pestana, João III-155
Pfister, Hans-Rüdiger III-577
Pinkwart, Niels IV-678
Pitakrat, Teerat I-527
Pizzolato, Ednaldo I-221
Plocher, Thomas III-346
Plocher, Tom III-329
Poirier, Franck II-214
Polzehl, Tim II-358
Ponnusamy, R. III-552, IV-274
Poor, G. Michael I-432, II-33, II-593
Popova, Severina II-23
Porras, Jari I-359

Qin, Hua IV-688
Qin, Yongqiang I-507
Quast, Holger III-418
Quax, Peter IV-668
Quiza, Phillip III-137

Rakiman, Kartini III-500
Rasooli, Amin IV-510
Rattanyu, Kanlaya III-586
Rau, Pei-Luen Patrick I-450, IV-688
Rauch, Thilo I-165
Ravaja, Niklas I-351, IV-466
Rayan, Infantdani A. II-523
Rebel, Matthias II-387
Reinkensmeyer, David IV-151
Ren, Jianfeng III-48
Rey, Beatriz I-423
Rhiu, Ilsun IV-257
Riahi, Ines III-92
Riahi, Meriem III-92
Richard, Paul II-140, II-158
Richardson, Kevin H. I-131
Riedel, Johann III-483
Riedenklau, Eckard II-110
Ritter, Helge II-110
Robert, Jean-Marc III-301
Robertson, Scott IV-167
Rodeiro Iglesias, Javier I-278, I-309
Roelands, Marc I-229
Roh, Yong-Wan II-464
Romero, Maximiliano IV-649
Ronzhin, Andrey II-454
Rösner, Dietmar I-121, II-396, III-593
Rothrock, Ling I-368, I-378, I-414
Rumeau, Pierre IV-530
Rusak, Zoltan II-204
Ryu, Taebeum IV-257

Saeed, Mehreen III-543
Saka, Emmanuel IV-285

Sakai, Akiko IV-413
Sakata, Nobuchika III-230
Sala, Pilar I-214
Salim, Siti Salwah IV-627
Samaras, George I-388
Sandnes, Frode Eika I-643, III-21
Santos, Caroline I-63
Sarker, Saonee IV-336
Sato, Makoto II-3
Saylor, Megan IV-580
Schels, Martin III-603
Scherer, Stefan III-603
Schlegel, Thomas I-288, I-618
Schlehuber, Christian III-205
Schmeier, Sven II-387
Schmidt, Miriam III-603
Schuller, Andreas I-268
Schwenker, Friedhelm III-603
Schwering, Angela II-71
Seals, Cheryl IV-3
Secore Levis, Melissa IV-266
Seissler, Marc I-165, I-299
Seo, Jonghoon I-637, I-653
Serna, Audrey IV-12
Shahid, Suleman III-500, III-543
Shi, Xiaoming IV-694
Shi, Yuanchun I-507, III-117, IV-368
Shiba, Haruya II-436
Shibuya, Yu III-101
Shigeno, Aguri II-378
Shih, Chi-Sheng III-107
Shih, Sheng-Cheng II-409
Shih, Ya-Chun IV-131
Shim, Soo In II-533
Shimamura, Kazunori II-436
Shimoda, Hiroshi III-465, IV-403
Shin, Hyunjin IV-484
Shin, Seungjae II-119
Shirai, Akira I-490
Shizuki, Buntarou II-81, II-557
Shneiderman, Ben I-3
Shovan, Curtis III-321
Siebert, Felix W. III-577
Siegel, Howard Jay I-450
Siegel, Marisa I-331
Simon, Lajos I-498
Sinnig, Daniel I-248
Son, Young-Jun I-368
Spath, Dieter I-268
Spies, Roland II-125
Stab, Christian I-480, IV-520
Stuerzlinger, Wolfgang II-331
Sun, Bo I-580
Sun, Xianghong III-346
Suo, Yue I-507
Suzuki, Akihiro IV-617
Suzuki, Shinya IV-237
Szilvasi, Lindsy III-500

Taghiyareh, Fattaneh IV-510
Takacs, Barnabas I-498
Takahashi, Tetsuro III-240
Takahashi, Yuzo II-314
Takao, Yuya II-603
Takeda, Hideaki III-271
Takeda, Hiroshi II-133
Tanabe, Naohisa IV-113
Tanaka, Jiro II-81, II-238, II-557
Tanaka, Kaori II-613
Tanaka, Sayaka IV-413
Tanaka, Yuya IV-318
Tang, Fengchun IV-336
Tang, Sheng Kai II-491
Taniguchi, Tadahiro I-461, I-490
Tariq, Hassan III-543
Tauber, Stefan IV-285
Teixeira-Faria, Pedro M. I-278, I-309
Tengku Wook, Tengku Siti Meriam IV-627
Terai, Hitoshi IV-384
Tews, Tessa-Karina III-577
Tharanathan, Anand I-378
Thiel, Simon III-474
Thiruvengada, Hari I-378
Tognetti, Simone IV-649
Tokumaru, Masataka IV-121
Tomimatsu, Kiyoshi I-45, III-30
Trapp, Marcus I-238
Trappeniers, Lieven I-229
Traue, Harald C. III-603
Troccaz, Jocelyne IV-202
Tsai, Chih-Chieh I-82
Tsai, Pei-Hsuan III-107
Tsai, Pen-Yan I-82
Tscherrig, Julien II-222
Tseng, Wen Chieh II-491
Tseng, Ya-Ying IV-570
Tsianos, Nikos I-388
Tsuchiya, Fumito I-588, I-628, IV-174, IV-210, IV-226

Tsujino, Yoshihiro II-351
Tullis, Tom I-331
Tulu, Bengisu IV-245
Tung, Fang-Wu IV-637
Turconi, Anna Carla I-406

Uang, Shiaw-Tsyr IV-490
Ullah, Sehat II-140
Ur, Blase III-137
Uster, Guillaume III-301

Valls, Alícia I-136
van Doesburg, Willem II-575
Vasconcelos, Verónica III-293
Vella, Frédéric IV-530
Vera-Muñoz, Cecilia III-129, IV-219
Vigouroux, Nadine IV-530
Vijayaraghavan, P. IV-274
Vladušič, Daniel III-519
Vogel-Heuser, Birgit IV-297

Wajima, Masayuki IV-617
Walter, Steffen III-603
Wan, Tao I-562
Wan Adnan, Wan Adilah I-17
Wang, Ai-Ling IV-107
Wang, Cheng-Han III-311
Wang, Hongmei III-338
Wang, Liang I-660
Wang, Mo I-142
Wang, Ying III-346
Wang, Yuxia IV-694
Waselewsky, Marc IV-678
Watanabe, Tomio II-585, II-603
Watanabe, Yoko IV-413
Wei, Ke III-220
Werner, Günter I-618
Wiley, Cyndi IV-285
Wohlfarter, Martin II-125
Wu, Andy II-148
Wu, Chenjun I-507
Wu, Hao I-580
Wu, Zhongke I-555, I-670
Wysk, Richard I-368

Xie, Bin I-580
Xie, Wenkui I-670
Xu, Qingqing I-677
Xu, Wenchang III-117

Yajima, Hiroshi IV-113
Yamaguchi, Takehiko II-3, II-62, II-158
Yamaguchi, Takumi II-436
Yamaguchi, Tatsuya IV-121
Yamaguchi, Yoshihisa III-357
Yamamoto, Michiya II-585, II-603
Yamamoto, Tomohito I-517
Yamanaka, Tsutomu IV-318
Yamanishi, Yuya II-436
Yang, Hsiao-Fang IV-77
Yang, Mau-Tsuen IV-131
Yasuda, Atsushi II-351
Yeh, Ching-Yu IV-570
Yen, Yi-Di III-185
Yin, Jing I-414
Yin, Qian I-598
Yin, Xiupu I-684
Yokokawa, Sho IV-237
Yokoyama, Saya II-436
Yoshimi, Hironori IV-226
Yoshino, Takashi II-368, II-378
Yu, Chun I-507
Yu, Guo-Jhen III-250
Yu, Lu III-346
Yun, Myung Hwan IV-257

Zablotskaya, Kseniya I-527
Zablotskiy, Sergey I-527
Zaki, Michael I-319
Zaragozá, Irene I-423
Zets, Gary III-321
Zhan, Jia-Xuan III-259
Zhang, Emily III-137
Zhang, Huiting III-346
Zhang, Kan III-346
Zhang, Shaopeng I-142
Zhang, Xian II-396
Zhang, Xuhui I-450
Zhao, Huiqin I-580
Zhao, Qingjie IV-694
Zheng, Xianjun Sam I-142
Zheng, Xin I-677
Zhou, Jinglin I-684
Zhou, Mingquan I-555, I-670
Zhou, Yun II-500
Zhu, Haijiaing I-684
Zhu, Miaoqi IV-659
Zimmerman, Guy W. I-432